PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark Office
Box PCT
Washington, D.C.20231
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 10 May 2000 (10.05.00)	
International application No. PCT/EP99/06081	**Applicant's or agent's file reference** EP-82 972/PC
International filing date (day/month/year) 19 August 1999 (19.08.99)	**Priority date** (day/month/year) 19 August 1998 (19.08.98)
Applicant BUSKIES, Christoph	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

17 March 2000 (17.03.00)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland**	Authorized officer Juan Cruz

PCT WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro

INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 7 :		(11) Internationale Veröffentlichungsnummer: **WO 00/11647**
G10L 13/06	A1	(43) Internationales Veröffentlichungsdatum: 2. März 2000 (02.03.00)

(21) Internationales Aktenzeichen: PCT/EP99/06081

(22) Internationales Anmeldedatum: 19. August 1999 (19.08.99)

(30) Prioritätsdaten:
198 37 661.8 19. August 1998 (19.08.98) DE

(71)(72) Anmelder und Erfinder: BUSKIES, Christoph [DE/DE]; Alsentrasse 21, D–22769 Hamburg (DE).

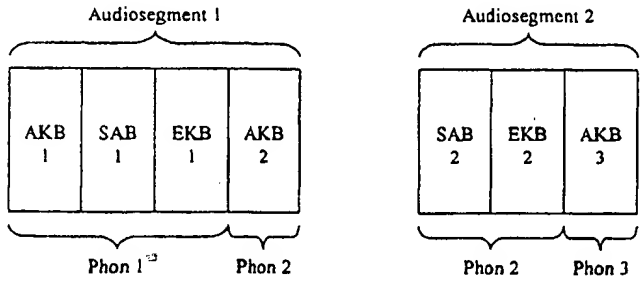
(74) Anwalt: SCHMIDT, Steffen, J.; Wuesthoff & Wuesthoff, Schweigerstrasse 2, D–81541 München (DE).

(81) Bestimmungsstaaten: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht
Mit internationalem Recherchenbericht.
Vor Ablauf der für Änderungen der Ansprüche zugelassenen Frist; Veröffentlichung wird wiederholt falls Änderungen eintreffen.

(54) Title: METHOD AND DEVICE FOR THE CONCATENATION OF AUDIOSEGMENTS, TAKING INTO ACCOUNT COARTICULATION

(54) Bezeichnung: VERFAHREN UND VORRICHTUNGEN ZUR KOARTIKULATIONSGERECHTEN KONKATENATION VON AUDIOSEGMENTEN



(57) Abstract

The invention makes it possible to synthesize any acoustic data by concatenation of individual audiosegment zones, the instant at which the concatenation of two successive audiosegment zones is carried out being chosen in accordance with properties of said audiosegments. In this manner synthesized acoustic data can be generated which after conversion into acoustic signals do not differ from the corresponding naturally produced acoustic signals. The invention notably makes it possible for synthesized speech data to be generated taking into account coarticulatory effects, by concatenation of individual speech–audiosegments. The speech data provided in this way can be converted into speech signals which are indistinguishable from natural spoken speech.

(57) Zusammenfassung

Die Erfindung ermöglicht es, beliebige akustische Daten durch eine Konkatenation einzelner Audiosegmentbereiche zu synthetisieren, wobei die Momente, zu denen die jeweilige Konkatenation zweier aufeinander folgender Audiosegmentbereiche erfolgt, in Abhängigkeit von Eigenschaften der Audiosegmente festgelegt werden. Auf diese Weise können synthetisierte akustische Daten erzeugt werden, die sich nach einer Umwandlung in akustische Signale nicht von entsprechenden natürlich erzeugten akustischen Signalen unterscheiden. Insbesondere erlaubt es die Erfindung, synthetisierte Sprachdaten unter Berücksichtigung koartikulatorischer Effekte durch Konkatenation einzelner Sprachaudiosegmente zu erzeugen. Die so zur Verfügung gestellten Sprachdaten können in Sprachsignale umgewandelt werden, die von einer natürlich gesprochenen Sprache nicht zuunterscheiden sind.

Verfahren und Vorrichtungen zur koartikulationsgerechten Konkatenation von Audiosegmenten

Die Erfindung betrifft ein Verfahren und eine Vorrichtung zur Konkatenation von Audiosegmenten zur Erzeugung synthetisierter akustischer Daten, insbesondere synthetisierter Sprache. Des weiteren betrifft die Erfindung synthetisierte Sprachsignale, die durch die erfindungsgemäße koartikulationsgerechte Konkatenation von Sprachsegmenten erzeugt wurden, sowie einen Datenträger, der ein Computerprogramm zur erfindungsgemäßen Erzeugung von synthetisierten akustischen Daten, insbesondere synthetisierter Sprache, enthält.

Zusätzlich betrifft die Erfindung einen Datenspeicher, der Audiosegmente enthält, die zur erfindungsgemäßen koartikulationsgerechten Konkatenation geeignet sind, und einen Tonträger, der erfindungsgemäß synthetisierte akustische Daten enthält.

Es ist zu betonen, daß sowohl der im folgenden dargestellte Stand der Technik als auch die vorliegenden Erfindung den gesamten Bereich der Synthese von akustischen Daten durch Konkatenation einzelner, auf beliebige Art und Weise erhaltene Audiosegmente betrifft. Aber um die Diskussion des Standes der Technik sowie die Beschreibung der vorliegenden Erfindung zu vereinfachen, beziehen sich die folgenden Ausführungen speziell auf synthetisierte Sprachdaten durch Konkatenation einzelner Sprachsegmente.

In den letzten Jahren hat sich im Bereich der Sprachsynthese der datenbasierte Ansatz gegenüber dem regelbasierten Ansatz durchgesetzt und ist in verschiedenen Verfahren und Systemen zur Sprachsynthese zu finden. Obwohl der regelbasierte Ansatz prinzipiell eine bessere Sprachsynthese ermöglicht, ist es für dessen Umsetzung notwendig, das gesamte zur Spracherzeugung notwendige Wissen explizit zu formulieren, d.h. die zu synthetisierende Sprache formal zu modellieren. Da die bekannten Sprachmodellierungen Vereinfachung der zu synthetisierenden Sprache aufweisen, ist die Sprachqualität der so erzeugten Sprache nicht ausreichend.

Daher wird in zunehmenden Maße eine datenbasierte Sprachsynthese durchgeführt, bei der aus einer einzelne Sprachsegmente aufweisenden Datenbasis entsprechende Segmente ausgewählt und miteinander verknüpft (konkateniert) werden. Die Sprachqualität hängt hierbei in erster Linie von der Zahl und Art der verfügbaren Sprachsegmente ab, denn es kann nur Sprache synthetisiert werden, die durch Sprachsegmente in der Datenbasis wiedergeben ist. Um die Zahl der vorzusehenden Sprachsegmente zu minimie-

ren und dennoch eine synthetisierte Sprache hoher Qualität zu erzeugen, sind verschieden Verfahren bekannt, die eine Verknüpfung (Konkatenation) der Sprachsegmente nach komplexen Regeln durchführen.

Unter Verwendung solcher Verfahren bzw. entsprechender Vorrichtungen kann ein Inventar, d.h. eine die Sprachaudiosegmente umfassende Datenbasis, verwendet werden, das vollständig und handhabbar ist. Ein Inventar ist vollständig, wenn damit jede Lautfolge der zu synthetisierenden Sprache erzeugt werden kann, und ist handhabbar, wenn die Zahl und Art der Daten des Inventars mit den technisch verfügbaren Mitteln in einer gewünschten Weise verarbeitet werden kann. Darüber hinaus muß ein solches Verfahren gewährleisten, daß die Konkatenation der einzelnen Inventarelemente eine synthetisierte Sprache erzeugt, die sich von einer natürlich gesprochenen Sprache möglichst wenig unterscheidet. Hierfür muß eine synthetisierte Sprache flüssig sein und die gleichen artikulatorischen Effekte einer natürlichen Sprache aufweisen. Hier kommen den sogenannten koartikulatorischen Effekten, d.h. der gegenseitigen Beeinflussung von Sprachlauten, eine besondere Bedeutung zu. Daher sollten die Inventarelemente so beschaffen sein, das sie die Koartikulation einzelner aufeinanderfolgender Sprachlaute berücksichtigen. Des weiteren sollte ein Verfahren zu Konkatenation der Inventarelemente, die Elemente unter Berücksichtigung der Koartikulation einzelner aufeinanderfolgender Sprachlaute sowie der übergeordneten Koartikulation mehrerer aufeinanderfolgender Sprachlaute, auch über Wort- und Satzgrenzen hinweg, verketten.

Vor der Darstellung des Standes der Technik werden im folgenden einige zum besseren Verständnis notwendige Begriffe aus dem Bereich der Sprachsynthese erläutert:

- Ein Laut ist eine Klasse von beliebigen Schallereignissen (Geräusche, Klänge, Töne usw). Die Schallereignisse werden gemäß eines Klassifikationsschemas in Lautklassen eingeteilt. Ein Schallereigniss gehört zu einem Laut, wenn hinsichtlich der zur Klassifikation verwendeten Parameter (z.B. Spektrum, Tonhöhe, Lautstärke, Brust- oder Kopfstimme, Koartikulation, Resonanzräume, Emotion usw.) die Werte des Schallereignis innerhalb der für den Laut definierten Wertebereiche liegen.
Das Klassifikationsschema für Laute hängt von der Art der Anwendung ab. Für Sprachlaute (= Phone) wird in der Regel die IPA-Klassifikation verwendet. Die hier verwendete Definition des Begriffes Laut ist jedoch nicht darauf beschränkt, sondern es lassen sich beliebige andere Parameter verwendet. Wird z.B. zusätzlich zu der IPA-Klassifikation noch die Tonhöhe oder der emotionale Ausdruck als Parameter in die Klassifikation mit einbezogen, so werden zwei 'a'-Laute mit unterschiedlicher Tonhöhe oder mit unter-

schiedlichem emotionalen Ausdruck zu unterschiedlichen Lauten im Sinne der Definition. Laute können aber auch die Töne eines Musikinstrumentes, etwa einer Geige, auf den unterschiedlichen Tonhöhen in den unterschiedlichen Spielweisen (Auf- und Abstrich, Detaché, Spiccato, Marcato, Pizzicato, col Legno etc.) sein. Laute können ebenso Hundegebell oder das Quietschen einer Autotüre sein.

Laute können durch Audiosegmente, die entsprechende akustische Daten enthalten, wiedergegeben werden.

In der auf die Definitionen folgenden Beschreibung der Erfindung kann immer der Begriff Phon durch den Begriff Laut im Sinne der vorigen Definition und der Begriff Phonem durch den Begriff Lautzeichen ersetzt werden. (Dies gilt auch umgekehrt, da Phone gemäß der IPA-Klassifikation eingeteilte Laute sind.)

- Ein statischer Laut hat Bereiche die ähnlich zu vorhergehenden oder nachfolgenden Bereichen des statischen Lauts sind. Die Ähnlichkeit muß nicht unbedingt eine exakte Entsprechung wie bei den Perioden eines Sinustones sein, sondern ist analog der Ähnlichkeit, die zwischen den Bereichen der unten definierten statischen Phone herrscht.

- Ein dynamischer Laut hat keine Bereiche, die vorhergenden oder nachfolgenden Bereichen des dynamischen Lautes ähneln, etwa das Schallereignis einer Explosion oder ein dynamisches Phon.

- Ein Phon ist ein von den Sprachorganen erzeugter Laut (ein Sprachlaut). Die Phone werden in statische und dynamische Phone unterteilt.

- Zu den statischen Phonen zählen Vokale, Diphtonge, Nasale, Laterale, Vibranten und Frikative.

- Zu den dynamischen Phonen zählen Plosive, Affrikate, Glottalstops und geschlagene Laute.

- Ein Phonem ist die formale Beschreibung eines Phons, wobei i. allg. die formale Beschreibung durch Lautschriftzeichen erfolgt.

- Die Koartikulation bezeichnet das Phänomen, daß ein Laut, also auch ein Phon, durch vorgelagerte und nachgelagerte Laute bzw. Phone beeinflußt wird, wobei die Koartikula-

tion sowohl zwischen unmittelbar benachbarten Lauten/Phonen auftritt, aber sich auch als übergeordnete Koartikulation über eine Folge mehrerer Laute/Phone erstrecken kann (Beispielsweise bei einer Lippenrundung).

Daher kann ein Laut bzw. Phon in drei Bereiche unterteilt werden (siehe auch Figur 1b):

- Der Anfangs-Koartikulationsbereich umfaßt den Bereich vom Beginn des Lautes/Phons bis zum Ende der Koartikulation aufgrund eines vorgelagerten Lautes/Phons.

- Der Soloartikulationsbereich, ist der Bereich des Lautes/Phons, der nicht durch einen vor- oder nachgelagerten Laut bzw. ein vor- oder nachgelagertes Phon beeinflußt ist.

- Der End-Koartikulationsbereich umfaßt den Bereich vom Beginn der Koartikulation aufgrund eines nachgelagerten Lautes/Phons bis zum Ende des Lautes/Phons.

- Der Koartikulationsbereich umfaßt einen End-Koartikulationsbereich und den benachbarten Anfangs-Koartikulationsbereich des benachbarten Lautes/Phons.

- Ein Polyphon ist eine Folge von Phonen.

- Die Elemente eines Inventars sind in kodierter Form gespeicherte Audiosegmente, die Laute, Teile von Lauten, Lautfolgen oder Teile von Lautfolgen, bzw. Phone, Teile von Phonen, Polyphone oder Teile von Polyphonen wiedergeben. Zur besseren Verständnis des möglichen Aufbau eines Audiosegmentes/Inventarelementes sei hier auf die Figur 2a, die ein herkömmliches Audiosegment zeigt, und die Figuren 2b-2l verwiesen, in denen erfindungsgemäße Audiosegmente gezeigt sind. Ergänzend ist zu erwähnen, daß Audiosegmente auch aus kleineren oder größeren Audiosegmenten gebildet werden können, die in dem Inventar oder einer Datenbank enthalten sind. Des weiteren können Audiosegmente auch in einer transformierten Form (z.B. einer fouriertransformierten Form) in dem Inventar oder einer Datenbank vorliegen. Audiosegmente für das vorliegende Verfahren können auch aus einem vorgelagerten Syntheseschritt (der nicht Teil des Verfahrens ist) stammen. Audiosegmente enthalten wenigstens einen Teil eines Anfangs-Koartikulationsbereiches, eines Soloartikulationsbereiches und/oder eines End-Koartikulationsbereiches. Anstelle von Audiosegmenten können auch Bereiche von Audiosegmenten verwendet werden.

- Unter Konkatenation versteht man das Aneinanderfügen zweier Audiosegmente.

- Der Konkatenationsmoment ist der Zeitpunkt, zu dem zwei Audiosegmente aneinandergefügt werden.

Die Konkatenation kann auf verschiedene Arten erfolgen, z.B. mit einem Crossfade oder einem Hardfade (siehe auch Figuren 3a-3e):

- Bei einem Crossfade werden ein zeitlich hinterer Bereich eines ersten Audiosegmentbereiches sowie ein zeitlich vorderer Bereich eines zweiten Audiosegmentbereiches mit geeigneten Übergangsfunktionen bearbeitet, und danach werden diese beiden Bereiche überlappend so addiert, daß maximal der zeitlich kürzere der beiden Bereichen von dem zeitlich längeren der beiden Bereiche vollständig überlappt wird.

- Bei einem Hardfade wird ein zeitlich hinterer Bereich eines ersten Audiosegmentes und ein zeitlich vorderer Bereich eines zweiten Audiosegmentes mit geeigneten Übergangsfunktionen bearbeitet, wobei diese beiden Audiosegmente so aneinandergefügt werden, daß sich der hintere Bereich des ersten Audiosegmentes und der vordere Bereich des zweiten Audiosegmentes nicht überlappen.

Der Koartikulationsbereich macht sich vor allem dadurch bemerkbar, daß eine Konkatenation darin mit Unstetigkeiten (z.B. Spektralsprüngen) verbunden ist.

Ergänzend sei zu erwähnen, daß streng genommen ein Hardfade einen Grenzfall eines Crossfades darstellt, bei dem eine Überlappung eines zeitlich hinteren Bereiches eines ersten Audiosegmentes und eines zeitlich vorderen Bereiches eines zweiten Audiosegmentes eine Länge Null hat. Dies erlaubt es in bestimmten, z.B. äußerst zeitkritischen Anwendungen einen Crossfade durch einen Hardfade zu ersetzen, wobei eine solche Vorgehensweise genau abzuwägen ist, da diese zu deutlichen Qualitätseinbußen bei der Konkatenation von Audiosegmenten führt, die eigentlich durch einen Crossfade zu konkatenieren sind.

- Unter Prosodie versteht man die Veränderungen der Sprachfrequenz und des Sprachrhythmus, die bei gesprochenen Worten bzw. Sätzen auftreten. Die Berücksichtigung solcher prosodischer Informationen ist bei der Sprachsynthese notwendig, um eine natürliche Wort- bzw. Satzmelodie zu erzeugen.

Aus WO 95/30193 ist ein Verfahren und eine Vorrichtung zur Umwandlung von Text in hörbare Sprachsignale unter Verwendung eines neuronalen Netzwerkes bekannt. Hierfür wird der in Sprache umzuwandelnde Text mit einer Konvertiereinheit in eine Folge von Phonemen umgewandelt, wobei zusätzlich Informationen über die syntaktischen Grenzen des Textes und die Betonung der einzelnen syntaktischen Komponenten des Textes erzeugt werden. Diese werden zusammen mit den Phonemen an eine Einrichtung weitergeleitet, die regelbasiert die Dauer der Aussprache der einzelnen Phoneme bestimmt. Ein Prozessor erzeugt aus jedem einzelnen Phonem in Verbindung mit den entsprechenden syntaktischen und zeitlichen Information eine geeignet Eingabe für das neuronale Netzwerk, wobei diese Eingabe für das neuronale Netz auch die entsprechenden prosodischen Informationen für die gesamte Phonemfolge umfaßt. Das neuronale Netz wählt aus den verfügbaren Audiosegmenten nun die aus, die die eingegebenen Phoneme am besten wiedergeben, und verkettet diese Audiosegmente entsprechend. Bei dieser Verkettung werden die einzelnen Audiosegmente in ihrer Dauer, Gesamtamplitude und Frequenz an vor- und nachgelagerte Audiosegmente unter Berücksichtigung der prosodischen Informationen der zu synthetisierenden Sprache angepaßt und zeitlich aufeinanderfolgend miteinander verbunden. Eine Veränderung einzelner Bereiche der Audiosegmente ist hier nicht beschrieben.

Zur Erzeugung der für dieses Verfahren erforderlichen Audiosegmente ist das neuronale Netzwerk zuerst zu trainieren, indem natürlich gesprochene Sprache in Phone oder Phonfolgen unterteilt wird und diesen Phonen oder Phonfolgen entsprechende Phonem oder Phonemfolgen in Form von Audiosegmenten zugeordnet werden. Da dieses Verfahren nur eine Veränderung von einzelnen Audiosegmenten, aber keine Veränderung einzelner Bereiche eines Audiosegmentes vorsieht, muß das neuronale Netzwerk mit möglichst vielen verschiedenen Phonen oder Phonfolgen trainiert werden, um beliebige Texte in synthetisierte natürlich klingende Sprache umzuwandeln. Dies kann sich je nach Anwendungsfall sehr aufwendig gestalten. Auf der anderen Seite kann ein unzureichender Trainingsprozeß des neuronalen Netzes die Qualität der zu synthetisierenden Sprache negativ beeinflussen. Des weiteren ist es bei dem hier beschriebene Verfahren nicht möglich, den Konkatenationsmoment der einzelnen Audiosegmente in Abhängigkeit vorgelagerter oder nachgelagerter Audiosegmente zu bestimmen, um so eine koartikulationsgerechte Konkatenation durchzuführen.

In US-5,524,172 ist eine Vorrichtung zur Erzeugung synthetisierter Sprache beschrieben, die das sogenannte Diphonverfahren nutzt. Hier wird ein Text, der in synthetisierte Sprache umgewandelt werden soll, in Phonemfolgen unterteilt, wobei jeder Phonemfolge ent-

sprechende prosodische Informationen zugeordnet werden. Aus einer Datenbank, die Audiosegmente in Form von Diphonen enthält, werden für jedes Phonem der Folge zwei das Phonem wiedergebende Diphone ausgewählt und unter Berücksichtigung der entsprechenden prosodischen Informationen konkateniert. Bei der Konkatenation werden die beiden Diphone jeweils mit Hilfe eines geeigneten Filters gewichtet und die Dauer und Tonhöhe beider Diphone so verändert, daß bei der Verkettung der Diphone eine synthetisierte Phonfolge erzeugt wird, deren Dauer und Tonhöhe der Dauer und Tonhöhe der gewünschten Phonemfolge entspricht. Bei der Konkatenation werden die einzelnen Diphone so addiert, daß sich ein zeitlich hinterer Bereich eines ersten Diphones und ein zeitlich vorderer Bereich eines zweiten Diphones überlappen, wobei der Konkatenationsmoment generell im Bereich stationären Bereiche der einzelnen Diphone liegt (siehe Figur 2a). Da eine Variation des Konkatenationsmomentes unter Berücksichtigung der Koartikulation aufeinanderfolgender Audiosegmente (Diphone) hier nicht vorgesehen ist, kann die Qualität (Natürlichkeit und Verständlichkeit) einer so synthetisierten Sprache negativ beeinflußt werden.

Eine Weiterentwicklung des zuvor diskutierten Verfahrens ist in EP-0,813,184 A1 zu finden. Auch hier wird ein in synthetisierte Sprache umzuwandelnder Text in einzelne Phoneme oder Phonemfolgen unterteilt und aus einer Datenbank entsprechende Audiosegmente ausgewählt und konkateniert. Um eine Verbesserung der synthetisierten Sprache zu erzielen, sind bei diesem Verfahren zwei Ansätze, die sich vom bisher diskutierten Stand der Technik unterscheiden, umgesetzt worden. Unter Verwendung eines Glättungsfilters, der die tieferfrequenten harmonischen Frequenzanteile eines vorgelagerten und eines nachgelagerten Audiosegments berücksichtigt, soll der Übergang von dem vorgelagerten Audiosegment zu dem nachgelagerten Audiosegment optimiert werden, indem ein zeitlich hinterer Bereich des vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des nachgelagerten Audiosegments im Frequenzbereich aufeinander abgestimmt werden. Des weiteren stellt die Datenbank Audiosegmente zur Verfügung, die sich leicht unterscheiden, aber zur Synthetisierung desselben Phonems geeignet sind. Auf diese Weise soll die natürliche Variation der Sprache nachgebildet werden, um eine höhere Qualität der synthetisierten Sprache zu erreichen. Sowohl die Verwendung des Glättungsfilter als auch die Auswahl aus einer Menge unterschiedlicher Audiosegmente zur Realisierung eines Phonems erfordert bei einer Umsetzung dieses Verfahrenes eine hohe Rechenleistung der verwendeten Systemkomponenten. Außerdem steigt der Umfang der Datenbank aufgrund der erhöhten Zahl der vorgesehenen Audiosegmente. Des weiteren ist auch bei diesem Verfahren eine koartikulationsabhängige Wahl

des Konkatenationsmoments einzelner Audiosegmente nicht vorgesehen, wodurch die Qualität der synthetisierten Sprache reduziert werden kann.

DE 693 18 209 T2 beschäftigt sich mit Formantsynthese. Gemäß diesem Dokument werden zwei mehrstimmige Laute unter Verwendung eines Interpolationsmechanismus miteinander verbunden, der auf ein letztes Phonem eines vorgelagerten Lauts und auf ein erstes Phonem eines nachgelagerten Lauts angewendet wird, wobei die zwei Phoneme der beiden Laute gleich sind und bei den verbundenen Lauten zu einem Phonem überlagert werden. Bei der Überlagerung werden die die zwei Phoneme beschreibenden Kurven jeweils mit einer Gewichtungsfunktion gewichtet. Die Gewichtungsfunktion wird bei jedem Phonem in einem Bereich angewendet, der unmittelbar nach dem Beginn des Phonems beginnt und unmittelbar vor dem Ende des Phonems endet. Somit entsprechen bei der hier beschriebenen Konkatenation von Lauten die verwendeten Bereiche der Phoneme, die den Übergang zwischen den Lauten bilden, im wesentlichen den jeweiligen gesamten Phonemen. Das heißt, daß die zur Konkatenation verwendeten Teile der Phoneme stets alle drei Bereiche, nämlich den jeweiligen Anfangs-Koartikulationsbereich, Soloartikulationsbereich und End-Koartikulationsbereich umfassen. Mithin lehrt D1 eine Verfahrensweise wie die Übergänge zwischen zwei Lauten zu glätten sind.

Des weiteren wird gemäß diesem Dokument der Moment der Konkatenation zweier Laute so festgelegt, daß sich das letzte Phonem in dem vorgelagerten Laut und das erste Phonem in dem nachgelagerten Laut vollständig überlappen.

Grundsätzlich ist festzustellen, daß DE 689 15 353 T2 eine Verbesserung der Tonqualität erreichen will indem eine Vorgehensweise angegeben wird, wie der Übergang zwischen zwei benachbarten Abtastwerten zu gestalten ist. Dies ist insbesondere bei niedrigen Abtastraten relevant.

Bei der in diesem Dokument beschriebenen Sprachsynthese werden Wellenformen verwendet, die zu konkatenierende Laute wiedergeben. Bei Wellenformen für vorgelagerte Laute wird jeweils ein entsprechender Endabtastwert und ein zugeordneter Nulldurchgangspunkt bestimmt, während bei Wellenformen für nachgelagerte Laute jeweils ein erster oberer Abtastwert und ein zugeordneter Nulldurchgangspunkt bestimmt wird. In Abhängigkeit dieser bestimmten Abtastwerte und der zugeordneten Nulldurchgangspunkte werden Laute auf maximal vier verschiedene Arten miteinander verbunden. Die Anzahl der Verbindungsarten wird auf zwei reduziert, wenn die Wellenformen unter Verwendung des Nyquist-Theorems erzeugt werden. In DE 689 15 353 T2 ist beschrieben,

daß sich der verwendete Bereich der Wellenformen zwischen dem letzten Abtastwert der vorgelagerten Wellenform und dem ersten Abtastwert der nachgelagerten Wellenform erstreckt. Eine Variation der Dauer der verwendeten Bereiche in Abhängigkeit der zu konkatenierenden Wellenformen, wie dies bei der Erfindung der Fall ist, ist in D1 nicht beschrieben.

Zusammenfassend ist zu sagen, daß es der Stand der Technik zwar erlaubt, beliebige Phonemfolgen zu synthetisieren, aber die so synthetisierten Phonemfolgen haben keine authentische Sprachqualität. Eine synthetisierte Phonemfolge hat eine authentische Sprachqualität, wenn sie von der gleichen Phonemfolge, die von einem realen Sprecher gesprochen wurde, durch einen Hörer nicht unterschieden werden kann.

Es sind auch Verfahren bekannt, die ein Inventar benutzen, das vollständige Worte und/oder Sätze in authentischer Sprachqualität als Inventarelemente enthält. Diese Elemente werden zur Sprachsynthese in einer gewünschten Reihenfolge hintereinander gesetzt, wobei die Möglichkeiten unterschiedliche Sprachsequenzen in hohem Maße von dem Umfang eines solchen Inventars limitiert werden. Die Synthese beliebiger Phonemfolgen ist mit diesen Verfahren nicht möglich.

Daher ist es eine Aufgabe der vorliegenden Erfindung ein Verfahren und eine entsprechende Vorrichtung zur Verfügung zu stellen, die die Probleme des Standes der Technik beseitigen und die Erzeugung synthetisierter akustischer Daten, insbesondere synthetisierter Sprachdaten, ermöglichen, die sich für einen Hörer nicht von entsprechenden natürlichen akustischen Daten, insbesondere natürlich gesprochener Sprache, unterscheiden. Die mit der Erfindung synthetisierten akustischen Daten, insbesondere synthetisierte Sprachdaten sollen eine authentische akustische Qualität, insbesondere eine authentische Sprachqualität aufweisen.

Zu Lösung dieser Aufgabe sieht die Erfindung ein Verfahren gemäß Anspruch 1, eine Vorrichtung gemäß Anspruch 14, synthetisierte Sprachsignale gemäß Anspruch 28, einen Datenträger gemäß Anspruch 39, einen Datenspeicher gemäß Anspruch 51, sowie einen Tonträger gemäß Anspruch 60 vor. Somit ermöglicht es die Erfindung, synthetisierte akustische Daten zu erzeugen, die eine Folge von Lauten wiedergeben, indem bei der Konkatenation von Audiosegmentbereichen der Moment der Konkatenation zweier Audiosegmentbereiche in Abhängigkeit von Eigenschaften der zu verknüpfenden Audiosegmentbereiche, insbesondere der die beiden Audiosegmentbereiche betreffenden Koartikulationseffekte bestimmt. Der Konkatenationsmoment wird gemäß der vorliegen-

den Erfindung vorzugsweise in der Umgebung der Grenzen des Solo-Artikulationsbereiches gewählt. Auf diese Weise wird eine Sprachqualität erreicht, die mit dem Stand der Technik nicht erzielbar ist. Dabei ist die erforderliche Rechenleistung nicht höher als beim Stand der Technik.

Um bei der Synthese akustischer Daten die Variationen nachzubilden, die bei entsprechenden natürlichen akustischen Daten zu finden sind, sieht die Erfindung eine unterschiedliche Auswahl der Audiosegmentbereiche sowie unterschiedliche Arten der koartikulationsgerechten Konkatenation vor. So wird ein höheres Maß an Natürlichkeit der synthetisierten akustischen Daten erzielt, wenn ein zeitlich nachgelagerter Audiosegmentbereich, dessen Anfang einen statischen Laut wiedergibt, mit einem zeitlich vorgelagerten Audiosegmentbereich mittels eines Crossfades verbunden wird, bzw. wenn ein zeitlich nachgelagerter Audiosegmentbereich, dessen Anfang einen dynamischen Laut wiedergibt, mit einem zeitlich vorgelagerten Audiosegmentbereich mittels eines Hardfades verbunden wird. Des weiteren ist es vorteilhaft den Anfang der zu erzeugenden synthetisierten akustischen Daten unter Verwendung eines den Anfang einer Lautfolge wiedergebenden Audiosegmentbereiches bzw. das Ende der zu erzeugenden synthetisierten akustischen Daten unter Verwendung eines das Ende einer Lautfolge wiedergebenden Audiosegmentbereiches zu erzeugen.

Um die Erzeugung der synthetisierten akustischen Daten einfacher und schneller durchzuführen, ermöglicht es die Erfindung die Zahl der zur Datensynthetisierung notwendigen Audiosegmentbereiche zu reduzieren, indem Audiosegmentbereiche verwendet werden, die immer mit der Wiedergabe eines dynamischen Lauts beginnen, wodurch alle Konkatenationen dieser Audiosegmentbereiche mittels eines Hardfades durchgeführt werden können. Hierfür werden zeitlich nachgelagerte Audiosegmentbereiche mit zeitlich vorgelagerten Audiosegmentbereichen verbunden, deren Anfänge jeweils einen dynamischen Laut wiedergeben. Auf diese Weise können auch mit geringer Rechenleistung (z.B. bei Anrufbeantwortern oder Autoleitsystemen) erfindungsgemäß synthetisierte akustische Daten hoher Qualität erzeugt werden.

Außerdem sieht die Erfindung vor, akustische Phänomene nachzubilden, die sich aufgrund einer gegenseitigen Beeinflussung einzelner Segmente entsprechender natürlicher akustischer Daten ergeben. Insbesondere ist hier vorgesehen, einzelne Audiosegmente bzw. einzelne Bereiche der Audiosegmente mit Hilfe geeigneter Funktionen zu bearbeiten. Somit kann u.a. die Frequenz, die Dauer, die Amplitude oder das Spektrum der Audiosegmente verändert werden. Werden mit der Erfindung synthetisierte Sprach-

daten erzeugt, so werden zur Lösung dieser Aufgabe vorzugsweise prosodische Informationen und/oder übergeordnete Koartikulationseffekte berücksichtigt.

Der Signalverlauf von synthetisierten akustischen Daten kann zusätzlich verbessert werden, wenn der Konkatenationsmoment an Stellen der einzelnen zu verknüpfenden Audiosegmentbereiche gelegt wird, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen. Diese Eigenschaften können u.a. sein: Nullstelle, Amplitudenwert, Steigung, Ableitung beliebigen Grades, Spektrum, Tonhöhe, Amplitudenwert in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften.

Darüber hinaus ermöglicht es Erfindung, die Auswahl der Audiosegmentbereiche zur Erzeugung der synthetisierten akustischen Daten zu verbessern sowie deren Konkatenation effizienter zu gestalten, indem heuristisches Wissen verwendet wird, das die Auswahl, Bearbeitung, Variation und Konkatenation der Audiosegmentbereiche betrifft.

Um synthetisierte akustische Daten zu erzeugen, die Sprachdaten sind, die sich von entsprechenden natürlichen Sprachdaten nicht unterscheiden, werden vorzugsweise Audiosegmentbereiche genutzt werden, die Laute/Phone oder Teile von Lautfolgen/Phonfolgen wiedergeben.

Außerdem erlaubt die Erfindung die Nutzung der erzeugten synthetisierten akustischen Daten, indem diese Daten in akustische Signale und/oder Sprachsignale umwandelbar und/ oder auf einem Datenträger speicherbar sind.

Des weiteren kann die Erfindung verwendet werden, um synthetisierte Sprachsignale zu Verfügung zu stellen, die sich von bekannten synthetisierten Sprachsignalen dadurch unterscheiden, daß sie sich in ihrer Natürlichkeit und Verständlichkeit nicht von realer Sprache unterscheiden. Hierfür werden Audiosegmentbereiche koartikulationsgerecht konkateniert, die jeweils Teile der Lautfolge/Phonfolge der zu synthetisierenden Sprache wiedergeben, indem die zu verwendenden Bereiche der Audiosegmente sowie der Moment der Konkatenation dieser Bereiche erfindungsgemäß wie in Anspruch 28 definiert bestimmt werden.

Eine zusätzliche Verbesserung der synthetisierten Sprache kann erreicht werden, wenn ein zeitlich nachgelagerter Audiosegmentbereich, dessen Anfang einen statischen Laut

bzw. ein statisches Phon wiedergibt, mit einem zeitlich vorgelagerten Audiosegmentbereich mittels eines Crossfades verbunden wird, bzw. wenn ein zeitlich nachgelagerter Audiosegmentbereich, dessen Anfang einen dynamischen Laut bzw. ein dynamisches Phon wiedergibt, mit einem zeitlich vorgelagerten Audiosegmentbereich mittels eines Hardfades verbunden wird. Hierbei umfassen statische Phone Vokale, Diphtonge, Liquide, Frikative, Vibranten und Nasale bzw. dynamische Phone Plosive, Affrikate, Glottalstops und geschlagene Laute.

Da sich die Anfangs- und Endbetonungen von Lauten bei natürlicher Sprache von vergleichbaren, aber eingebetteten Lauten unterscheiden, ist es zu bevorzugen, entsprechende Audiosegmentbereiche zu verwenden, deren Anfänge jeweils den Anfang bzw. deren Enden jeweils das Ende von zu synthetisierender Sprache wiedergeben.

Besonders bei Erzeugung synthetisierter Sprache ist eine schnelle und effiziente Vorgehensweise wünschenswert. Hierfür ist es zu bevorzugen, erfindungsgemäße koartikulationsgerechte Konkatenationen immer mittels Hardfades durchzuführen, wobei nur Audiosegmentbereiche verwendet werden, deren Anfänge jeweils immer einen dynamischen Laut bzw. ein dynamisches Phon wiedergeben. Derartige Audiosegmentbereiche können mit der Erfindung durch koartikulationsgerechte Konkatenation entsprechender Audiosegmentbereiche zuvor erzeugt werden.

Des weiteren stellt die Erfindung Sprachsignale bereit, die einen natürlichen Sprachfluß, Sprachmelodie und Sprachrhythmus haben, indem Audiosegmentbereiche jeweils vor und/oder nach der Konkatenation in ihrer Gesamtheit oder in einzelnen Bereichen mit Hilfe geeigneter Funktionen bearbeitet werden. Besonders vorteilhaft ist es diese Variation zusätzlich in Bereichen durchzuführen, in denen die entsprechenden Momente der Konkatenationen liegen, um u.a. die Frequenz, Dauer, Amplitude oder das Spektrum zu verändern.

Ein zusätzlich verbesserter Signalverlauf kann erreicht werden, wenn die Konkatenationsmomente an Stellen der zu verknüpfenden Audiosegmentbereiche liegen, an denen diese in einer oder mehrerer geeigneter Eigenschaften übereinstimmen.

Um eine einfache Nutzung und/oder Weiterverarbeitung der erfindungsgemäßen Sprachsignale durch bekannte Verfahren oder Vorrichtungen, z.B. einem CD-Abspielgerät, zu erlauben, ist es besonders zu bevorzugen, daß die Sprachsignale in akustische Signale umwandelbar oder auf einem Datenträger speicherbar sind.

Um die Erfindung auch bei bekannten Vorrichtungen, z.B. einem Personal Computer oder einem computergesteuerten Musikinstrument, anzuwenden, ist ein Datenträger vorgesehen, der ein Computerprogramm enthält, der die Durchführung des erfindungsgemäßen Verfahrens bzw. die Steuerung der erfindungsgemäßen Vorrichtung sowie deren verschiedenen Ausführungsformen ermöglicht. Des weiteren erlaubt der erfindungsgemäße Datenträger auch die Erzeugung von Sprachsignalen, die koartikulationsgerechte Konkatenationen aufweisen.

Um ein Audiosegmente umfassendes Inventar zur Verfügung zu stellen, mit dem synthetisierte akustische Daten, insbesondere synthetisierte Sprachdaten, erzeugt werden können, die sich von entsprechenden natürlichen akustischen Daten nicht unterscheiden, sieht die Erfindung einen Datenspeicher vor, der Audiosegmente enthält, die geeignet sind, um erfindungsgemäß zu synthetisierten akustischen Daten konkateniert zu werden. Vorzugsweise enthält ein solcher Datenträger Audiosegmente, zur Durchführung des erfindungsgemäßen Verfahrens, zur Anwendung bei der erfindungsgemäßen Vorrichtung oder dem erfindungsgemäßen Datenträger geeignet sind. Alternativ kann der Datenträger auch erfindungsgemäße Sprachsignale umfassen.

Darüber hinaus ermöglicht es die Erfindung, erfindungsgemäße synthetisierte akustische Daten, insbesondere synthetisierte Sprachdaten, zur Verfügung zu stellen, die mit herkömmlichen bekannten Vorrichtungen, beispielsweise einem Tonbandgerät, einem CD-Spieler oder einer PC-Audiokarte, genutzt werden können. Hierfür ist ein Tonträger vorgesehen, der Daten aufweist, die zumindest teilweise mit dem erfindungsgemäßen Verfahren oder der erfindungsgemäßen Vorrichtung bzw. unter Verwendung des erfindungsgemäßen Datenträgers oder des erfindungsgemäßen Datenspeichers erzeugt wurden, Der Tonträger kann auch Daten enthalten, die erfindungsgemäß koartikulationsgerecht konkatenierte Sprachsignale sind.

Weitere Eigenschaften, Merkmale, Vorteile oder Abwandlungen der Erfindung werden anhand der nachfolgenden Beschreibung erläutert. Dabei zeigt:

Figur 1a: Schematische Darstellung einer erfindungsgemäßen Vorrichtung zur Erzeugung synthetisierter akustischer Daten;

Figur 1b: Struktur eines Lautes / Phons.

Figur 2a: Struktur eines herkömmlichen Audiosegmentes nach dem Stand der Technik, aus Teilen von zwei Lauten bestehend, also ein Diphon für Sprache. Wesentlich ist, daß die Solo-Artikulations-Bereiche jeweils nur teilweise im herkömmlichen Diphon-Audiosegment enthalten sind.

Figur 2b: Struktur eines erfindungsgemäßen Audiosegmentes, das Teile eines Lautes/Phons mit nachgelagerten Koartikulationsbereichen (für Sprache quasi ein 'verschobenes' Diphon) wiedergibt.

Figur 2c: Struktur eines erfindungsgemäßen Audiosegmentes, das Teile eines Lautes/Phons mit vorgelagerten Koartikulationsbereichen wiedergibt.

Figur 2d: Struktur eines erfindungsgemäßen Audiosegmentes, das Teile eines Lautes/Phons mit nachgelagerten Koartikulationsbereichen wiedergibt und zusätzliche Bereiche enthält.

Figur 2e: Struktur eines erfindungsgemäßen Audiosegmentes, das Teile eines Lautes/Phons mit vorgelagerten Koartikulationsbereichen wiedergibt und zusätzliche Bereiche enthält.

Figur 2f: Struktur eines erfindungsgemäßen Audiosegmentes, das Teile mehrerer Laute/Phone (für Sprache: ein Polyphon) mit jeweils nachgelagerten Koartikulationsbereichen wiedergibt. Die Laute / Phone 2 bis (n-1) sind jeweils vollständig im Audiosegment enthalten.

Figur 2g: Struktur eines erfindungsgemäßen Audiosegmentes, das Teile mehrerer Laute/Phone (für Sprache: ein Polyphon) mit jeweils vorgelagerten Koartikulationsbereichen wiedergibt. Die Laute / Phone 2 bis (n-1) sind jeweils vollständig im Audiosegment enthalten.

Figur 2h: Struktur eines erfindungsgemäßen Audiosegmentes, das Teile mehrerer Laute/Phone (für Sprache: ein Polyphon) mit jeweils nachgelagerten Koartikulationsbereichen wiedergibt und zusätzliche Bereiche enthält. Die Laute / Phone 2 bis (n-1) sind jeweils vollständig im Audiosegment enthalten.

Figur 2i: Struktur eines erfindungsgemäßen Audiosegmentes, das Teile mehrerer Laute/Phone (für Sprache: ein Polyphon) mit jeweils vorgelagerten Koartikulationsbereichen

wiedergibt und zusätzliche Bereiche enthält. Die Laute / Phone 2 bis (n-1) sind jeweils vollständig im Audiosegment enthalten.

Figur 2j: Struktur eines erfindungsgemäßen Audiosegmentes, das einen Teil eines Lautes / Phons vom Anfang einer Lautfolge / Phonfolge wiedergibt.

Figur 2k: Struktur eines erfindungsgemäßen Audiosegmentes, das Teile von Lauten / Phonens vom Anfang einer Lautfolge / Phonfolge wiedergibt.

Figur 2l: Struktur eines erfindungsgemäßen Audiosegmentes, das einen Laut / ein Phon vom Ende einer Lautfolge / Phonfolge wiedergibt.

Figur 3a: Konkatenation gemäß dem Stand der Technik am Beispiel zweier herkömmlicher Audiosegmente. Die Segmente beginnen und enden mit Teilen der Solo-Aktikulationsbereiche (in der Regel jeweils die Hälfte).

Figur 3al: Konkatenation gemäß dem Stand der Technik. Der Solo-Artikulationsbereich des mittleren Phons stammt aus zwei unterschiedlichen Audiosegmenten.

Figur 3b: Konkatenation nach dem erfindungsgemäßen Verfahren am Beispiel zweier Audiosegmente, die je einen Laut / ein Phon mit nachgelagerten Koartikulationsbereichen enthalten. Beide Laute / Phone stammen aus der Mitte einer Lauteinheitenfolge

Figur 3bl: Konkatenation dieser Audiosegmente mittels eines Crossfades. Der Soloartikulationsbereich stammt aus einem Audiosegment. Der Übergang zwischen den Audiosegmenten erfolgt zwischen zwei Bereichen und ist somit unempfindlicher gegen Unterschiede (im Spektrum, Frequenz, Amplitude etc.). Die Audiosegmente können vor der Konkatenation auch noch mit zusätzlichen Übergangsfunktionen bearbeitet werden.

Figur 3bll: Konkatenation dieser Audiosegmente mittels eines Hardfades.

Figur 3c: Konkatenation gemäß dem erfindungsgemäßen Verfahren am Beispiel zweier erfindungsgemäßer Audiosegmente, die je einen Laut / ein Phon mit nachgelagerten Koartikulationsbereichen enthalten, wobei das erste Audiosegment vom Anfang einer Lautfolge stammt.

Figur 3cI: Konkatenation dieser Audiosegmente mittels eines Crossfades.

Figur 3cII: Konkatenation dieser Audiosegmente mittels eines Hardfades.

Figur 3d: Konkatenation gemäß dem erfindungsgemäßen Verfahren am Beispiel zweier erfindungsgemäßer Audiosegmente, die je einen Laut / ein Phon mit vorgelagerten Koartikulationsbereichen enthalten. Beide Audiosegmente stammen aus der Mitte einer Lautfolge.

Figur 3dI: Konkatenation dieser Audiosegmente mittels eines Crossfades. Der Soloartikulationsbereich stammt aus einem Audiosegment.

Figur 3dII: Konkatenation dieser Audiosegmente mittels eines Hardfades.

Figur 3e: Konkatenation nach dem erfindungsgemäßen Verfahren am Beispiel zweier erfindungsgemäßer Audiosegmente, die je einen Laut / ein Phon mit nachgelagerten Koartikulationsbereichen enthalten, wobei das letzte Audiosegment vom Ende einer Lautfolge stammt.

Figur 3eI: Konkatenation dieser Audiosegmente mittels eines Crossfades.

Figur 3eII: Konkatenation dieser Audiosegmente mittels eines Hardfades.

Figur 4: Schematische Darstellung der Schritte eines erfindungsgemäßen Verfahrens zur Erzeugung synthetisierter akustischer Daten.

Die im folgenden benutzten Bezugszeichen beziehen sich auf die Figur 1a und die im folgenden für die verschiedenen Verfahrensschritte benutzten Nummern beziehen sich auf die Figur 4.

Um mit Hilfe der Erfindung beispielsweise einen Text in synthetisierte Sprache umzuwandeln, ist es notwendig in einem vorgelagerten Schritt diesen Text in eine Folge von Lautzeichen bzw. Phonemen unter Verwendung bekannter Verfahren oder Vorrichtungen zu unterteilen. Vorzugsweise sind auch dem Text entsprechende prosodische Informationen zu erzeugen. Die Lautfolge bzw. Phonfolge sowie die prosodischen und zusätzlichen Informationen dienen als Eingabegrößen für das erfindungsgemäße Verfahren bzw. die erfindungsgemäße Vorrichtung.

Die zu synthetisierenden Laute/Phone werden einer Eingabeeinheit 101 der Vorrichtung 1 zur Erzeugung synthetisierter Sprachdaten zugeführt und in einer ersten Speichereinheit 103 abgelegt (siehe Figur 1a). Mit Hilfe einer Auswahleinrichtung 105 werden aus einem Audiosegmente (Elemente) enthaltenden Inventar, das in einer Datenbank 107 gespeichert ist, oder von einer vorgeschalteten Syntheseeinrichtung 108 (die nicht Bestandteil der Erfindung ist) die Audiosegmentbereiche ausgewählt, die Laute bzw. Phone oder Teile von Lauten bzw. Phonen wiedergeben, die den einzelnen eingegebenen Lautzeichen bzw. Phonemen oder Teilen davon entsprechen und in einer Reihenfolge, die der Reihenfolge der eingegebenen Lautzeichen bzw. Phoneme entspricht, in einer zweiten Speichereinheit 109 gespeichert. Falls das Inventar Teile von Lautfolgen oder von Polyphonen wiedergebende Audiosegmente enthält, so wählt die Auswahleinrichtung 105 vorzugsweise die Audiosegmente aus, die die meisten Teile von Lautfolgen bzw. von Polyphonen wiedergeben, die einer Folge von Lautzeichen bzw. Phonemen aus der eingegebenen Lautzeichenfolge bzw. Phonemfolge entsprechen, so daß eine minimale Anzahl von Audiosegmenten zur Synthese der eingegebenen Phonemfolge benötigt wird.

Stellt die Datenbank 107 oder die vorgeschaltete Syntheseeinrichtung 108 ein Inventar mit Audiosegmenten unterschiedlicher Arten zur Verfügung, so wählt die Auswahleinrichtung 105 vorzugsweise die längsten Audiosegmentbereiche aus, dieTeile der Lautfolge/Phonfolge wiedergeben , um die eingegebene Lautfolge bzw. Phonfolge und/oder eine Folge von Lauten/ Phonen aus einer minimalen Anzahl von Audiosegmentbereichen zu synthetisieren. Hierbei ist es vorteilhaft, verkettete Laute/Phone wiedergebende Audiosegmentbereiche zu verwenden, die einen zeitlich vorgelagerten statischen Laut/Phon und einen zeitlich nachgelagerten dynamischen Laut/Phon wiedergeben. So entstehen Audiosegmente, die aufgrund der Einbettung der dynamischen Laute/Phone immer mit einem statischen Laut/Phon beginnen. Dadurch vereinfacht und vereinheitlicht sich das Vorgehen bei Konkatenationen solcher Audiosegmente, da hierfür nur Crossfades benötigt werden.

Um eine koartikulationsgerechte Konkatenation der zu verkettenden Audiosegmentbereiche zu erzielen, werden mit Hilfe einer Konkatenationseinrichtung 111 die Konkatenationsmomente zweier aufeinanderfolgender Audiosegmentbereiche wie folgt festgelegt:

- Soll ein Audiosegmentbereich zu Synthetisierung des Anfanges der eingegebenen Lautfolge/Phonfolge (Schritt 1) verwendet werden, so ist aus dem Inventar ein Audio-

segmentbereich zu wählen, das den Anfang einer Lautfolge/Phonfolge wiedergibt und mit einem zeitlich nachgelagerten Audiosegmentbereich zu verketten (siehe Figur 3c und Schritt 3 in Figur 4).

- Bei der Konkatenation eines zweiten Audiosegmentbereiches an einen zeitlich vorgelagerten ersten Audiosegmentbereich ist zu unterscheiden, ob der zweite Audiosegmentbereich mit der Wiedergabe eines statischen Lautes/Phons oder eines dynamischen Lautes/Phons beginnt, um die Wahl des Momentes der Konkatenation entsprechend zu treffen (Schritt 6).

- Beginnt der zweite Audiosegmentbereich mit einem statischen Laut/Phon, wird die Konkatenation in Form eines Crossfades durchgeführt, wobei der Moment der Konkatenation im zeitlich hinteren Bereich des ersten Audiosegmentbereiches und im zeitlich vorderen Bereich des zweiten Audiosegmentbereiches gelegt wird, wodurch sich diese beiden Bereiche bei der Konkatenation überlappen oder wenigstens unmittelbar aneinandergrenzen (siehe Figuren 3bI, 3cI, 3dI und 3eI, Konkatenation mittels Crossfade).

- Beginnt der zweite Audiosegmentbereich mit einem dynamischen Laut/Phon, wird die Konkatenation in Form eines Hardfades durchgeführt, wobei der Moment der Konkatenation zeitlich unmittelbar hinter der zeitlich hinteren Bereich des ersten Audiosegmentbereiches und zeitlich unmittelbar vor dem zeitlich vorderen Bereich des zweiten Audiosegmentbreiches gelegt wird (siehe Figuren 3bII, 3cII, 3dII und 3eII, Konkatenation mittels Hardfade).

Auf diese Weise können aus diesen ursprünglich verfügbaren Audiosegmentbereichen neue Audiosegmente erzeugt werden, die mit der Wiedergabe eines statischen Lautes/Phons beginnen. Dies erreicht man, indem Audiosegmentbereiche, die mit der Wiedergabe eines dynamischen Lautes/Phons beginnen, zeitlich nachgelagert mit Audiosegmentbereichen, die mit der Wiedergabe eines statischen Lautes/Phons beginnen, verkettet werden. Dies vergrößert zwar die Zahl der Audiosegmente bzw. den Umfang des Inventars, kann aber bei der Erzeugung synthetisierter Sprachdaten einen rechentechnischen Vorteil darstellen, da weniger einzelne Konkatenationen zur Erzeugung einer Lautfolge/Phonemfolge erforderliche sind und Konkatenationen nur noch in Form eines Crossfades durchgeführt werden müssen. Vorzugsweise werden die so erzeugten neuen verketteten Audiosegmente der Datenbank 107 oder einer anderen Speichereinheit 113 zugeführt.

Ein weiterer Vorteil dieser Verkettung der ursprüngliche Audiosegmentbereiche zu neuen längeren Audiosegmenten ergibt sich, wenn sich beispielsweise eine Folge von Lauten/Phonen in der eingegebenen Lautfolge/Phonfolge häufig wiederholt. Dann kann auf eines der neuen entsprechend verketteten Audiosegmente zurückgegriffen werden und es ist nicht notwendig, bei jedem Auftreten dieser Folge von Lauten/Phonen eine erneute Konkatenation der ursprünglich vorhandenen Audiosegmentbereiche durchzuführen. Vorzugsweise sind bei der Speicherung solcher verketteten Audiosegmente auch übergreifende Koartikulationseffekte zu erfassen bzw. spezifische Koartikulationseffekte in Form zusätzlicher Daten dem gespeicherten verketteten Audiosegment zuzuordnen.

Soll ein Audiosegmentbereich zu Synthetisierung des Endes der eingegebenen Lautfolge/Phonfolge verwendet werden, so ist aus dem Inventar ein Audiosegmentbereich zu wählen, das ein Ende einer Lautfolge/Phonfolge wiedergibt und mit einem zeitlich vorgelagerten Audiosegmentbereich zu verketten (siehe Figur 3e und Schritt 8 in Figur 4).

Die einzelnen Audiosegmente werden in der Datenbank 107 kodiert gespeichert, wobei die kodierte Form der Audiosegmente neben der Wellenform des jeweiligen Audiosegmentes angeben kann, welche Teile von Lautfolgen/Phonfolgen das jeweilige Audiosegment wiedergibt, welche Art der Konkatenation (z.B. Hardfade, linearer oder exponentieller Crossfade) mit welchem zeitlich nachfolgenden Audiosegmentbereich durchzuführen ist und zu welchem Moment die Konkatenation mit welchem zeitlich nachfolgenden Audiosegmentbereich stattfindet. Vorzugsweise enthält die kodierte Form der Audiosegmente auch Informationen bezüglich der Prosodie, übergeordneten Koartikulationen und Übergangsfunktionen, die verwendet werden, um eine zusätzliche Verbesserung der Sprachqualität zu erzielen.

Bei der Wahl der Audiosegmentbereiche zur Synthetisierung der eingegebenen Lautfolge/Phonfolge werden als zeitlich nachgelagerte Audiosegmentbereiche solche gewählt, die den Eigenschaften der jeweils zeitlich vorgelagerten Audiosegmentbereiche, u.a. Konkatenationsart und Konkatenationsmoment, entsprechen. Nachdem die jeweils Teile der Lautfolge/Phonfolge wiedergebenden Audiosegmentbereiche aus der Datenbank 107 oder der vorgeschalteten Syntheseeinrichtung 108 gewählt wurden, erfolgt die Verkettung zweier aufeinanderfolgender Audiosegmentbereiche mit Hilfe der Konkatenationseinrichtung 111 folgendermaßen. Es wird die Wellenform, die Konkatenationsart, der Konkatenationsmoment sowie evtl. zusätzliche Informationen des ersten Audiosegmentbereiches und des zweiten Audiosegmentbereiches aus der Datenbank oder der Syntheseeinrichtung (Figur 3b und Schritt 10 und 11) geladen. Vorzugsweise werden bei der

oben erwähnten Wahl der Audiosegmentbereiche solche Audiosegmentbereiche gewählt, die hinsichtlich ihrer Konkatenationsart und ihres Konkatenationsmoments zu einander passen. In diesem Fall ist das Laden der Informationen bezüglich der Konkatenationsart und des Konkatenationsmomentes des zweiten Audiosegmentbereiches nicht mehr notwendig.

Zur Konkatenation der beiden Audiosegmentbereiche werden die Wellenform des ersten Audiosegmentbereiches in einem zeitlich hinteren Bereich und die Wellenform des zweiten Audiosegmentbereiches in einem zeitlich vorderen Bereich jeweils mit geeigneten Übergangsfunktionen bearbeitet, z.B. mit einer geeigneten Gewichtungsfunktion multipliziert (siehe Figur 3b, Schritt 12 und 13). Die Längen des zeitlich hinteren Bereiches des ersten Audiosegmentbereiches und des zeitlich vorderen Bereiches des zweiten Audiosegmentbereiches ergeben sich aus der Konkatenationsart und zeitlichen Lage des Konkatenationsmomentes, wobei diese Längen auch in der kodierten Form der Audiosegmente in der Datenbank gespeichert werden können.

Sind die beiden Audiosegmentbereiche mit einem Crossfade zu verketten, werden diese entsprechend dem jeweiligen Konkatenationsmoment überlappend addiert (siehe Figuren 3bI, 3cI, 3dI und 3eI, Schritt 15). Vorzugsweise ist hierbei ein linearer symmetrischer Crossfade zu verwenden, es kann aber auch jede andere Art eines Crossfades oder jede Art von Übergangsfunktionen eingesetzt werden. Ist eine Konkatenation in Form eines Hardfades durchzuführen, werden die beiden Audiosegmentbereiche nicht überlappend hintereinander verbunden (siehe Figur 3bII, 3cII, 3dII und 3eII, Schritt 15). Wie in Figur 3bII zu sehen ist, werden hierbei die beiden Audiosegmentbereiche zeitlich unmittelbar hintereinander angeordnet. Um die so erzeugten synthetisierten Sprachdaten weiterverarbeiten zu können, werden diese vorzugsweise in einer dritten Speichereinheit 115 abgelegt.

Für die weitere Verkettung mit nachfolgenden Audiosegmentbereichen werden die bisher verketteten Audiosegmentbereiche als erster Audiosegmentbereich betrachtet (Schritt 16) und der oben beschriebenen Verkettungsprozeß solange wiederholt, bis die gesamte Lautfolge/Phonfolge synthetisiert wurde.

Zur Verbesserung der Qualität der synthetisierten Sprachdaten sind vorzugsweise auch die prosodischen und zusätzlichen Informationen, die zusätzlich zu der Lautfolge/Phonfolge eingegeben werden, bei der Verkettung der Audiosegmentbereiche zu berücksichtigen. Mit Hilfe bekannter Verfahren kann die Frequenz, Dauer, Amplitude

und/oder spektralen Eigenschaften der Audiosegmentbereiche vor und/oder nach deren Konkatenation so verändert werden, daß die synthetisierten Sprachdaten eine natürliche Wort- und/oder Satzmelodie aufweisen (Schritte 14, 17 oder 18). Hierbei ist es zu bevorzugen, Konkatenationsmomente an Stellen der Audiosegmentbereiche zu wählen, an denen diese in einer oder mehrerer geeigneter Eigenschaften übereinstimmen.

Um die Übergänge zwischen zwei aufeinander folgenden Audiosegmentbereichen zu optimieren, ist zusätzlich die Bearbeitung der beiden Audiosegmentbereiche mit Hilfe geeigneter Funktionen im Bereich des Konkatenationsmomentes vorgesehen, um u.a. die Frequenzen, Dauern, Amplituden und spektralen Eigenschaften anzupassen. Des weiteren erlaubt es die Erfindung, auch übergeordnete akustische Phänomene einer realen Sprache, wie z.B. übergeordnete Koartikulationseffekte oder Sprachstil (u.a. Flüstern, Betonung, Gesangsstimme, Falsett, emotionaler Ausdruck) bei der Synthetisierung der Lautfolge/Phonfolgen zu berücksichtigen. Hierfür werden Informationen, die solche übergeordnete Phänomene betreffen, zusätzlich in kodierter Form mit den entsprechenden Audiosegmenten gespeichert, um so bei der Auswahl der Audiosegmentbereiche nur solche zu wählen, die den übergeordneten Koartikulationseigenschaften der zeitlich vor- und/oder nachgelagerten Audiosegmentbereichen entsprechen.

Die so erzeugten synthetisierten Sprachdaten haben vorzugsweise eine Form, die es unter Verwendung einer Ausgabeeinheit 117 erlaubt, die Sprachdaten in akustische Sprachsignale umzuwandeln und die Sprachdaten und/oder Sprachsignale auf einem akustischen, optischen, magnetischen oder elektrischen Datenträger zu speichern (Schritt 19).

Im allgemeinen werden Inventarelemente durch die Aufnahme von real gesprochener Sprache erzeugt. In Abhängigkeit des Trainingsgrades des inventaraufbauenden Sprechers, d.h. seiner Fähigkeit die aufzunehmende Sprache zu kontrollieren (z.B. die Tonhöhe der Sprache zu kontrollieren oder exakt auf einer Tonhöhe zu sprechen), ist es möglich, gleiche oder ähnliche Inventarelemente zu erzeugen, die verschobene Grenzen zwischen den Solo-Artikulationsbereichen und Koartikulationsbereichen haben. Dadurch ergeben sich wesentlich mehr Möglichkeiten, die Konkatenationspunkte an verschiedenen Stellen zu plazieren. In der Folge kann die Qualität einer zu synthetisierenden Sprache deutlich verbessert werden.

Mit dieser Erfindung ist es erstmals möglich synthetisierte Sprachsignale durch eine koartikulationsgerechte Konkatenation einzelner Audiosegmentbereiche zu erzeugen, da

der Moment der Konkatenation in Abhängigkeit der jeweils zu verkettenden Audiosegmentbereiche gewählt wird. Auf diese Weise kann eine synthetisierte Sprache erzeugt werden, die vom einer natürlichen Sprache nicht mehr zu unterscheiden ist. Im Gegensatz zu bekannten Verfahren oder Vorrichtungen werden die hier verwendeten Audiosegmente nicht durch ein Einsprechen ganzer Worte erzeugt, um eine authentische Sprachqualität zu gewährleisten. Daher ist es mit dieser Erfindung möglich, synthetisierte Sprache beliebigen Inhalts in der Qualität einer real gesprochenen Sprache zu erzeugen.

Obwohl diese Erfindung am Beispiel der Sprachsynthese beschrieben wurde, ist die Erfindung nicht auf den Bereich der synthetisierten Sprache beschränkt, sondern kann zu Synthetisierung beliebiger akustischer Daten, bzw. beliebiger Schallereignisse verwendet werden. Daher ist diese Erfindung auch für eine Erzeugung und/oder Bereitstellung von synthetisierten Sprachdaten und/oder Sprachsignale für beliebige Sprachen oder Dialekte sowie auch zur Synthese von Musik einsetzbar.

Ansprüche

1. Verfahren zur koartikulationsgerechten Konkatenation von Audiosegmenten, um synthetisierte akustische Daten zu erzeugen, die eine Folge konkatenierter Laute wiedergeben, mit folgenden Schritten:
- Auswahl von wenigstens zwei Audiosegmenten, die Bereiche enthalten, die jeweils einen Teil eines Lautes oder einen Teil der Lautfolge wiedergeben, aufweist,
gekennzeichnet durch die Schritte:
- Festlegen eines zu verwendenden Bereiches eines zeitlich vorgelagerten Audiosegments,
- Festlegen eines zu verwendenden Bereiches eines zeitlich nachgelagerten Audiosegments, der zeitlich unmittelbar vor dem zu verwendenden Bereich des zeitlich nachgelagerten Audiosegments beginnt und mit dem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich des zeitlich nachgelagerten Audiosegments endet,
- wobei die Dauer und Lage der zu verwendenden Bereiche in Abhängigkeit der vor- und nachgelagerten Audiosegmente bestimmt wird, und
- Konkatenieren des festgelegten Bereiches des zeitlich vorgelagerten Audiosegments mit dem festgelegten Bereich des zeitlich nachgelagerten Audiosegments, indem der Moment der Konkatenation in Abhängigkeit von Eigenschaften des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments in dessen festgelegten Bereich gelegt wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß
- der Moment der Konkatenation in einen Bereich gelegt wird, der in der Umgebung der Grenzen des zuerst zu verwendenden Soloartikulationsbereichs des zeitlich nachgelagerten Audiosegments liegt, wenn dessen zu verwendender Bereich am Anfang einen statischen Laut wiedergibt, und
- ein zeitlich hinterer Bereich des zu verwendenden Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des zu verwendenden Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und überlappend addiert werden (Crossfade), wobei die Übergangsfunktionen und die Länge eines Überlappungsbereichs der beiden Bereiche in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt werden.

3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß

- der Moment der Konkatenation in einen Bereich gelegt wird, der zeitlich unmittelbar vor dem zu verwendenden Bereich des zeitlich nachgelagerten Audiosegments liegt, wenn dessen verwendeter Bereich am Anfang einen dynamischen Laut wiedergibt, und
- ein zeitlich hinterer Bereich des zu verwendenden Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des zu verwendenden Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und nicht überlappend verbunden werden (Hardfade), wobei die Übergangsfunktionen in Abhängigkeit der zu synthetisierenden akustischen Daten bestimmt werden.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß für einen Laut oder einen Teil der Folge konkatenierter Laute am Anfang der konkatenierten Lautfolge ein Bereich eines Audiosegmentes ausgewählt wird, so daß der Anfang des Bereiches die Eigenschaften des Anfangs der konkatenierten Lautfolge wiedergibt.

5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß für einen Laut oder einen Teil der Folge konkatenierter Laute am Ende der konkatenierten Lautfolge ein Bereich eines Audiosegmentes ausgewählt wird, so daß das Ende des Bereiches die Eigenschaften des Endes der konkatenierten Lautfolge wiedergibt.

6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß die zu synthetisierenden Sprachdaten in Gruppen zusammengefaßt werden, die jeweils durch ein einzelnes Audiosegment beschrieben werden.

7. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, daß für den zeitlich nachgelagerten Audiosegmentbereich ein Audiosegmentbereich gewählt wird, der die größte Anzahl aufeinanderfolgender Teile der Laute der Lautfolge wiedergibt, um bei der Erzeugung der synthetisierten akustischen Daten die kleinste Anzahl von Audiosegmentbereichen zu verwenden.

8. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der konkatenierten Lautfolge durchgeführt wird, wobei diese Eigenschaften u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein können.

9. Verfahren nach einem der Ansprüche 1 bis 8, dadurch gekennzeichnet, daß

eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in einem Bereich durchgeführt wird, in dem der Moment der Konkatenation liegt. Dies kann u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein.

10. Verfahren nach einem der Ansprüche 1 bis 9, dadurch gekennzeichnet, daß der Moment der Konkatenation an Stellen in den zu verwendenden Bereichen des zeitlich vorgelagerten und/oder des zeitlich nachgelagerten Audiosegments gelegt wird, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen, wobei diese Eigenschaften u.a. sein können: Nullstelle, Amplitudenwert, Steigung, Ableitung beliebigen Grades, Spektrum, Tonhöhe, Amplitudenwert in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften.

11. Verfahren nach einem der Ansprüche 1 bis 10, dadurch gekennzeichnet, daß
- die Auswahl der verwendeten Bereiche einzelner Audiosegmente, deren Bearbeitung, deren Variation sowie deren Konkatenation zusätzlich unter Verwendung heuristischen Wissens durchgeführt wird, das durch ein zusätzlich durchgeführtes heuristisches Verfahren gewonnen wird.

12. Verfahren einem der Ansprüche 1 bis 11, dadurch gekennzeichnet, daß
- die zu synthetisierenden akustischen Daten Sprachdaten und die Laute Phone sind,
- die statischen Laute Vokale, Diphtonge, Liquide, Vibranten, Frikative und Nasale umfassen, und
- die dynamischen Laute Plosive, Affrikate, Glottalstops und geschlagenen Laute umfassen.

13. Verfahren nach einem der Ansprüche 1 bis 12, dadurch gekennzeichnet, daß
- eine Umwandlung der synthetisierten akustischen Daten in akustische Signale und/oder Sprachsignale durchgeführt wird.

14. Vorrichtung zur koartikulationsgerechten Konkatenation von Audiosegmenten, um synthetisierte akustische Daten zu erzeugen, die eine Folge von Lauten wiedergeben, mit:
- einer Datenbank, in der Audiosegmente gespeichert sind, die jeweils Teile eines Lautes oder Teile einer Folge von (konkatenierten) Lauten wiedergeben

- und/oder einer beliebigen vorgeschalteten Syntheseeinrichtung (nicht Bestandteil dieser Erfindung), die Audiosegmente liefert, - einer Einrichtung zur Auswahl von wenigstens zwei Audiosegmenten aus der Datenbank und/oder der vorgeschalteten Syntheseeinrichtung, und
- einer Einrichtung zur Konkatenation der Audiosegmente, dadurch gekennzeichnet, daß die Konkatenationseinrichtung geeignet ist,
- einen zu verwendenden Bereiches eines zeitlich vorgelagerten Audiosegments zu definieren,
- einen zu verwendenden Bereiches eines zeitlich nachgelagerten Audiosegments in einem Bereich zu definieren, der mit dem zeitlich nachgelagerten Audiosegment beginnt und zeitlich nach einem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich des zeitlich nachgelagerten Audiosegmentes endet,
- die Dauer und Lage der verwendeten Bereiche in Abhängigkeit der vor- und nachgelagerten Audiosegmente zu bestimmen, und
- den verwendeten Bereich des zeitlich vorgelagerten Audiosegments mit dem verwendeten Bereich des zeitlich nachgelagerten Audiosegments durch Definition des Moment der Konkatenation in Abhängigkeit von Eigenschaften des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments in einem Bereich zu konkatenieren, der zeitlich unmittelbar vor dem verwendeten Bereich des zeitlich nachgelagerten Audiosegments beginnt und mit dem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich des zeitlich nachgelagerten Audiosegments endet.

15. Vorrichtung nach Anspruch 14, dadurch gekennzeichnet, daß die Konkatenationeinrichtung umfaßt:
- Einrichtungen zur Konkatenation des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments mit dem verwendeten Bereich des zeitlich nachgelagerten Audiosegment, dessen verwendeter Bereich am Anfang einen statischen Laut wiedergibt, in der Umgebung der Grenzen des zuerst auftretetenden Soloartikulationsbereichs des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments,
- Einrichtungen zur Bearbeitung eines zeitlich hinteren Bereiches des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und eines zeitlich vorderen Bereiches des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen, und
- Einrichtungen zur überlappenden Addition der beiden Bereiche in einem von den zu konkatenierenden Audiosegmenten abhängenden Überlappungsbereich (Crossfade), wobei die Übergangsfunktionen und die Länge eines Überlappungsbereiches der beiden Bereiche in Abhängigkeit der zu synthetisierenden akustischen Daten bestimmt werden.

16. Vorrichtung nach Anspruch 14, dadurch gekennzeichnet, daß die Konkatenationseinrichtung umfaßt:
- Einrichtungen zur Konkatenation des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments mit dem verwendeten Bereich des zeitlich nachgelagerten Audiosegment, dessen verwendeter Bereich am Anfang einen dynamischen Laut wiedergibt, zeitlich unmittelbar vor dem verwendeten Bereich des zeitlich nachgelagerten Audiosegments,
- Einrichtungen zur Bearbeitung eines zeitlich hinteren Bereiches des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und eines zeitlich vorderen Bereiches des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen, wobei die Übergangsfunktionen in Abhängigkeit der zu synthetisierenden akustischen Daten bestimmt werden, und
- Einrichtungen zur nicht überlappenden Verbindung der Audiosegmente.

17. Vorrichtung nach einem der Ansprüche 14 bis 16, dadurch gekennzeichnet, daß die Datenbank Audiosegmente enthält oder die vorgeschaltete Syntheseeinrichtung Audiosegmente liefert, die Bereiche enthalten, die zu Beginn einen Laut oder einen Teil der konkatenierten Lautfolge am Anfang der konkatenierten Lautfolge wiedergibt.

18. Vorrichtung nach einem der Ansprüche 14 bis 17, dadurch gekennzeichnet, daß die Datenbank Audiosegmente enthält oder die vorgeschaltete Syntheseeinrichtung Audiosegmente liefert, die Bereiche enthalten, deren Ende einen Laut oder einen Teil der konkatenierten Lautfolge am Ende der konkatenierten Lautfolge wiedergibt.

19. Vorrichtung nach einem der Ansprüche 14 bis 18, dadurch gekennzeichnet, daß die Datenbank eine Gruppe von Audiosegmenten enthält oder die vorgeschaltete Syntheseeinrichtung Audiosegmente liefert, die Bereiche enthalten, deren Anfänge jeweils nur einen statischen Laut wiedergeben.

20. Vorrichtung nach einem der Ansprüche 14 bis 19, dadurch gekennzeichnet, daß die Konkatenationseinrichtung umfaßt:
- Einrichtungen zur Erzeugung weiterer Audiosegmente durch Konkatenation von Bereichen von Audiosegmenten, wobei die Anfänge der Bereiche jeweils einen statischen Laut wiedergeben, jeweils mit einem Bereich eines zeitlich nachgelagerten Audiosegment, dessen verwendeter Bereich am Anfang einen dynamischen Laut wiedergibt, und

- eine Einrichtung, die die weiteren Audiosegmente der Datenbank oder der Auswahleinrichtung zuführt.

21. Vorrichtung nach einem der Ansprüche 14 bis 20, dadurch gekennzeichnet, daß die Auswahleinrichtung geeignet ist, bei der Auswahl der Audiosegmentbereiche aus der Datenbank oder der vorgeschalteten Syntheseeinrichtung, die Audiosegmentbereiche auszuwählen, die jeweils die meisten aufeinanderfolgenden Teile der konkatenierten Laute der konkatenierten Lautfolge wiedergeben.

22. Vorrichtung nach einem der Ansprüche 14 bis 21, dadurch gekennzeichnet, daß die Konkatenationseinrichtung Einrichtungen zur Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der konkatenierten Lautfolge aufweist. Dies kann u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein.

23. Vorrichtung nach einem der Ansprüche 14 bis 22, dadurch gekennzeichnet, daß
- die Konkatenationseinrichtung Einrichtungen zur Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in einem den Moment der Konkatenation umfassenden Bereich aufweist, wobei diese Funktion u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein kann.

24. Vorrichtung nach einem der Ansprüche 14 bis 23, dadurch gekennzeichnet, daß
- die Konkatenationseinrichtung Einrichtungen zur Auswahl des Momentes der Konkatenation bei einer Stelle in den verwendeten Bereichen des zeitlich vorgelagerten und/oder des zeitlich nachgelagerten Audiosegments, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen, wobei diese Eigenschaften u.a. sein können: Nullstelle, Amplitudenwert, Steigung, Ableitung beliebigen Grades, Spektrum, Tonhöhe, Amplitudenwert in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften.

25. Vorrichtung nach einem der Ansprüche 14 bis 24, dadurch gekennzeichnet, daß
- die Auswahleinrichtung Einrichtungen zur Implementation heuristischen Wissens umfaßt, das die Auswahl der verwendeten Bereiche der einzelnen Audiosegmente, deren Bearbeitung, deren Variation sowie deren Konkatenation betrifft.

26. Vorrichtung nach einem der Ansprüche 14 bis 25, dadurch gekennzeichnet, daß

- die Datenbank Audiosegmente enthält oder die vorgeschaltete Syntheseeinrichtung Audiosegmente liefert, die Bereiche enthalten, die jeweils wenigstens einen Teil eines Lautes bzw. Phons, einen Laut bzw. ein Phon, Teile von Lautfolgen bzw. Polyphonen oder Lautfolgen bzw. Polyphone wiedergeben, wobei ein statischer Laut einen statischen Phon entspricht und Vokale, Diphtonge, Liquide, Vibranten, Frikative und Nasale umfaßt und
ein dynamischer Laut einem dynamischen Phon entspricht und Plosive, Affrikate, Glottalstops und geschlagene Laute umfaßt, und
- die Konkatenationseinrichtung geeignet ist, um durch Konkatenation von Audiosegmenten synthetisierte Sprachdaten zu erzeugen.

27. Vorrichtung nach einem der Ansprüche 14 bis 26, dadurch gekennzeichnet, daß
- Einrichtungen zur Umwandlung der synthetisierten akustischen Daten in akustische Signale und/oder Sprachsignale vorhanden sind.

28. Synthetisierte Sprachsignale, die aus einer Folge von Lauten bzw. Phonen bestehen, wobei die Sprachsignale erzeugt werden, indem:
- wenigstens zwei die Laute bzw. Phone wiedergebende Audiosegmente ausgewählt werden, und
- die Audiosegmente durch eine koartikulationsgerechte Konkatenation verkettet werden, wobei
- ein zu verwendender Bereich eines zeitlich vorgelagerten Audiosegments festgelegt wird,
- ein zu verwendender Bereich eines zeitlich nachgelagerten Audiosegments festgelegt wird, der zeitlich unmittelbar vor dem zu verwendenden Bereich des zeitlich nachgelagerten Audiosegments beginnt und mit dem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich des zeitlich nachgelagerten Audiosegments endet,
- wobei die Dauer und Lage der zu verwendenden Bereiche in Abhängigkeit der Audiosegmente bestimmt wird, und
- die verwendeten Bereiche der Audiosegmente koartikulationsgerecht konkateniert werden, indem der Moment der Konkatenation in Abhängigkeit von Eigenschaften des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments in dessen festgelegten Bereich gelegt wird.

29. Synthetisierte Sprachsignale nach Anspruch 28, dadurch gekennzeichnet, daß die Sprachsignale erzeugt werden, indem

- die Audiosegmente zu einem Moment konkateniert werden, der in der Umgebung der Grenzen des zuerst auftretenden Soloartikulationsbereichs des verwendeten Bereiches des zeitlich nachgelagerten Audiosegmentes liegt, wenn der Anfang dieses Bereiches einen statischen Laut bzw. ein statisches Phon wiedergibt , wobei ein statisches Phon ein Vokal, ein Diphtong, ein Liquid, ein Frikativ, ein Vibrant oder ein Nasal ist, und
- ein zeitlich hinterer Bereich des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und beide Bereiche überlappend addiert werden (Crossfade), wobei die Übergangsfunktionen und die Länge eines Überlappungsbereichs beiden Bereiche in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt werden.

30. Synthetisierte Sprachsignale nach Anspruch 28, dadurch gekennzeichnet, daß die Sprachsignale erzeugt werden, indem
- die Audiosegmente zu einem Moment konkateniert werden, der zeitlich unmittelbar vor dem verwendeten Bereich des zeitlich nachgelagerten Audiosegmentes liegt, wenn der Anfang dieses Bereiches einen dynamischen Laut bzw. ein dynamisches Phon wiedergibt, wobei ein dynamisches Phon ein Plosiv, ein Affrikat, ein Glottalstop oder ein geschlagener Laut ist, und
- ein zeitlich hinterer Bereich des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet werden und nicht überlappend verbunden werden (Hardfade) wobei die Übergangsfunktionen in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt werden.

31. Synthetisierte Sprachsignale nach einem der Ansprüche 28 bis 30, dadurch gekennzeichnet, daß
- der erste Laut bzw. das erste Phon oder ein Teil der ersten Lautfolge bzw. des ersten Polyphons in der Folge durch ein Audiosegment erzeugt wird, dessen verwendeter Bereich am Anfang die Eigenschaften des Anfangs der Folge wiedergibt.

32. Synthetisierte Sprachsignale nach einem der Ansprüche 28 bis 31, dadurch gekennzeichnet, daß
- der letzte Laut bzw. das letzte Phon oder ein Teil der letzten Lautfolge bzw. des letzten Polyphon in der Folge durch ein Audiosegment erzeugt wird, dessen verwendeter Bereich am Ende die Eigenschaften des Endes der Folge wiedergibt.

33. Synthetisierte Sprachsignale nach einem der Ansprüche 28 bis 32, dadurch gekennzeichnet, daß

- die Sprachsignale erzeugt werden indem nachgelagerte mit der Wiedergabe eines dynamischen Lautes bzw. Phons beginnenden Bereiche von Audiosegmenten mit vorgelagerten mit der Wiedergabe eines statischen Lautes bzw. Phons beginnende Bereichen von Audiosegmenten konkateniert werden.

34. Synthetisierte Sprachsignale nach einem der Ansprüche 28 bis 33, dadurch gekennzeichnet, daß

- die Audiosegmentbereiche ausgewählt werden, die die meisten Teile von Lauten bzw. Phonen der Folge wiedergeben, um bei der Erzeugung der Sprachsignale die minimale Anzahl von Audiosegmentbereichen zu verwenden.

35. Synthetisierte Sprachsignale nach einem der Ansprüche 28 bis 34, dadurch gekennzeichnet, daß

- die Sprachsignale durch Konkatenation der verwendeten Bereiche von Audiosegmenten erzeugt werden, die mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der Lautfolge bzw. Phonfolge bearbeitet werden. Dies kann u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein.

36. Synthetisierte Sprachsignale einem der Ansprüche 28 bis 35, dadurch gekennzeichnet, daß

- die Sprachsignale durch Konkatenation der verwendeten Bereiche von Audiosegmenten erzeugt werden, die mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der Lautfolge bzw. Phonfolge in einem Bereich bearbeitet werden, in dem der Moment der Konkatenation liegt, wobei diese Eigenschaften u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein können.

37. Synthetisierte Sprachsignale einem der Ansprüche 28 bis 36, dadurch gekennzeichnet, daß der Moment der Konkatenation bei einer Stelle in den verwendeten Bereichen des vorgelagerten und/oder des nachgelagerten Audiosegmentes liegt, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen, wobei diese Eigenschaften u.a. sein können: Nullstelle, Amplitudenwert, Steigung, Ableitung beliebigen Grades, Spektrum, Tonhöhe, Amplitudenwert in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften.

38. Synthetisierte Sprachsignale nach einem der Ansprüche 28 bis 37, dadurch gekennzeichnet, daß die Sprachsignale geeignet sind, in akustische Signale umgewandelt zu werden.

39. Datenträger, der ein Computerprogramm zur koartikulationsgerechten Konkatenation von Audiosegmenten enthält, um synthetisierte akustische Daten zu erzeugen, die eine Folge konkatenierter Laute wiedergeben, mit folgenden Schritten:
- Auswahl von wenigstens zwei Audiosegmenten, die Bereiche enthalten, die jeweils einen Teil eines Lautes oder einen Teil der Folge konkatenierter Laute wiedergeben,
gekennzeichnet durch die Schritte:
- Festlegen eines zu verwendenden Bereiches eines zeitlich vorgelagerten Audiosegments,
- Festlegen eines zu verwendenden Bereiches eines zeitlich nachgelagerten Audiosegments, der zeitlich unmittelbar vor dem zu verwendenden Bereich des zeitlich nachgelagerten Audiosegments beginnt und mit dem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich des zeitlich nachgelagerten Audiosegments endet,
- wobei die Dauer und Lage der zu verwendenden Bereiche in Abhängigkeit der vor- und nachgelagerten Audiosegmente bestimmt wird, und
- Konkatenieren des festgelegten Bereiches des zeitlich vorgelagerten Audiosegments mit dem festgelegten Bereich des zeitlich nachgelagerten Audiosegments, indem der Moment der Konkatenation in Abhängigkeit von Eigenschaften des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments in dessen festgelegten Bereich gelegt wird.

40. Datenträger nach Anspruch 39, dadurch gekennzeichnet, daß das Computerprogramm den Moment der Konkatenation des verwendeten Bereiches des zweiten Audiosegmentes mit dem verwendeten Bereich des ersten Audiosegment so wählt, daß
- der Moment der Konkatenation in einen Bereich gelegt wird, der in der Umgebung der Grenzen des zuerst verwendeten Soloartikulationsbereichs des zeitlich nachgelagerten Audiosegments liegt, wenn dessen verwendeter Bereich am Anfang einen statischen Laut wiedergibt, und
- ein zeitlich hinterer Bereich des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und überlappend addiert werden (Crossfade), wobei Übergangsfunktionen und die Länge

eines Überlappungsbereichs der beiden Bereiche in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt wird.

41. Datenträger nach Anspruch 39 dadurch gekennzeichnet, daß das Computerprogramm den Moment der Konkatenation des verwendeten Bereiches des zweiten Audiosegmentes mit dem verwendeten Bereich des ersten Audiosegmentes so wählt, daß
- der Moment der Konkatenation in einen Bereich gelegt wird, der zeitlich unmittelbar vor dem verwendeten Bereich des zeitlich nachgelagerten Audiosegments liegt, wenn dessen verwendeter Bereich am Anfang einen dynamischen Laut wiedergibt, und
- ein zeitlich hinterer Bereich des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und nicht überlappend verbunden werden (Hardfade), wobei die Übergangsfunktionen in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt werden.

42. Datenträger nach einem der Ansprüche 39 bis 41, dadurch gekennzeichnet, daß das Computerprogramm für einen Laut oder einen Teil der Folge konkatenierter Laute am Anfang der konkatenierten Lautfolge einen Bereich eines Audiosegments auswählt, dessen Anfang die Eigenschaften des Anfangs der konkatenierten Lautfolge wiedergibt.

43. Datenträger nach einem der Ansprüche 39 bis 42, dadurch gekennzeichnet, daß das Computerprogramm für einen Laut oder einen Teil der Folge konkatenierter Laute am Ende der konkatenierten Lautfolge einen Bereich eines Audiosegments auswählt, dessen Ende die Eigenschaften des Endes der konkatenierten Lautfolge wiedergibt.

44. Datenträger nach einem der Ansprüche 39 bis 43, dadurch gekennzeichnet, daß das Computerprogramm eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der Lautfolge durchführt. Dies kann u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein.

45. Datenträger nach einem der Ansprüche 39 bis 44, dadurch gekennzeichnet, daß das Computerprogramm für den zeitlich nachgelagerten Audiosegmentbereich einen Audiosegmentbereich wählt, der die größte Anzahl aufeinanderfolgender Teile der konkatenierter Laute der Lautfolge wiedergibt, um bei der Erzeugung der synthetisierten akustischen Daten die kleinste Anzahl von Audiosegmentbereichen zu verwenden.

46. Datenträger nach einem der Ansprüche 39 bis 45, dadurch gekennzeichnet, daß das Computerprogramm eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in einem Bereich durchführt, in dem der Moment der Konkatenation liegt. Dies kann u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein.

47. Datenträger nach einem der Ansprüche 39 bis 46, dadurch gekennzeichnet, daß Computerprogramm den Moment der Konkatenation bei einer Stelle in den verwendeten Bereichen des ersten und/oder des zweiten Audiosegmentes festlegt, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen, wobei diese Eigenschaften u.a. sein können: Nullstelle, Amplitudenwert, Steigung, Ableitung beliebigen Grades, Spektrum, Tonhöhe, Amplitudenwert in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften.

48. Datenträger nach einem der Ansprüche 39 bis 47, dadurch gekennzeichnet, daß das Computerprogramm eine Implementation von heuristischem Wissen durchführt, das die Auswahl der verwendeten Bereiche der einzelnen Audiosegmente, deren Bearbeitung, deren Variation sowie deren Konkatenation betrifft.

49. Datenträger nach einem der Ansprüche 39 bis 48, dadurch gekennzeichnet, daß das Computerprogramm zur Erzeugung synthetisierter Sprachdaten geeignet ist, wobei die Laute Phone sind, die statischen Laute Vokale, Diphtonge, Liquide, Vibranten, Frikative und Nasale und die dynamischen Laute Plosive, Affrikate, Glottalstops und geschlagene Laute umfassen.

50. Datenträger nach einem der Ansprüche 39 bis 49, dadurch gekennzeichnet, daß das Computerprogramm die synthetisierten akustischen Daten in akustische umwandelbare Daten und/oder Sprachsignale umwandelt.

51. Akustischer, optischer, magnetischer oder elektrischer Datenspeicher, der Audiosegmente enthält, um durch eine Konkatenation von verwendeten Bereichen der Audiosegmente unter Verwendung des Verfahrens nach Anspruch 1 oder der Vorrichtung nach Anspruch 14 oder des Datenträgers nach Anspruch 39 synthetisierte akustische Daten zu erzeugen.

52. Datenspeicher nach Anspruch 51, dadurch gekennzeichnet, daß eine Gruppe der Audiosegmente Laute bzw. Phone oder Teile von Lauten bzw. Phonen wiedergeben.

53. Datenspeicher nach Anspruch 51 oder 52, dadurch gekennzeichnet, daß eine Gruppe der Audiosegmente Lautfolgen oder Teile von Lautfolgen bzw. Polyphone oder Teile von Polyphonen wiedergeben.

54. Datenspeicher nach einem der Ansprüche 50 bis 53, dadurch gekennzeichnet, daß eine Gruppe von Audiosegmenten zur Verfügung gestellt wird, deren verwendete Bereiche mit einem statischen Laut bzw. Phon beginnen, wobei die statischen Phone Vokale, Diphtonge, Liquide, Frikative, Vibranten und Nasale umfassen.

55. Datenspeicher nach einem der Ansprüche 50 bis 54, dadurch gekennzeichnet, daß Audiosegmente zur Verfügung gestellt werden, die geeignet sind in akustische Signale umgewandelt zu werden.

56. Datenspeicher nach einem der Ansprüche 50 bis 55, der zusätzlich Informationen enthält, um eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der zu synthetisierenden akustischen Daten durchzuführen. Dies kann u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein.

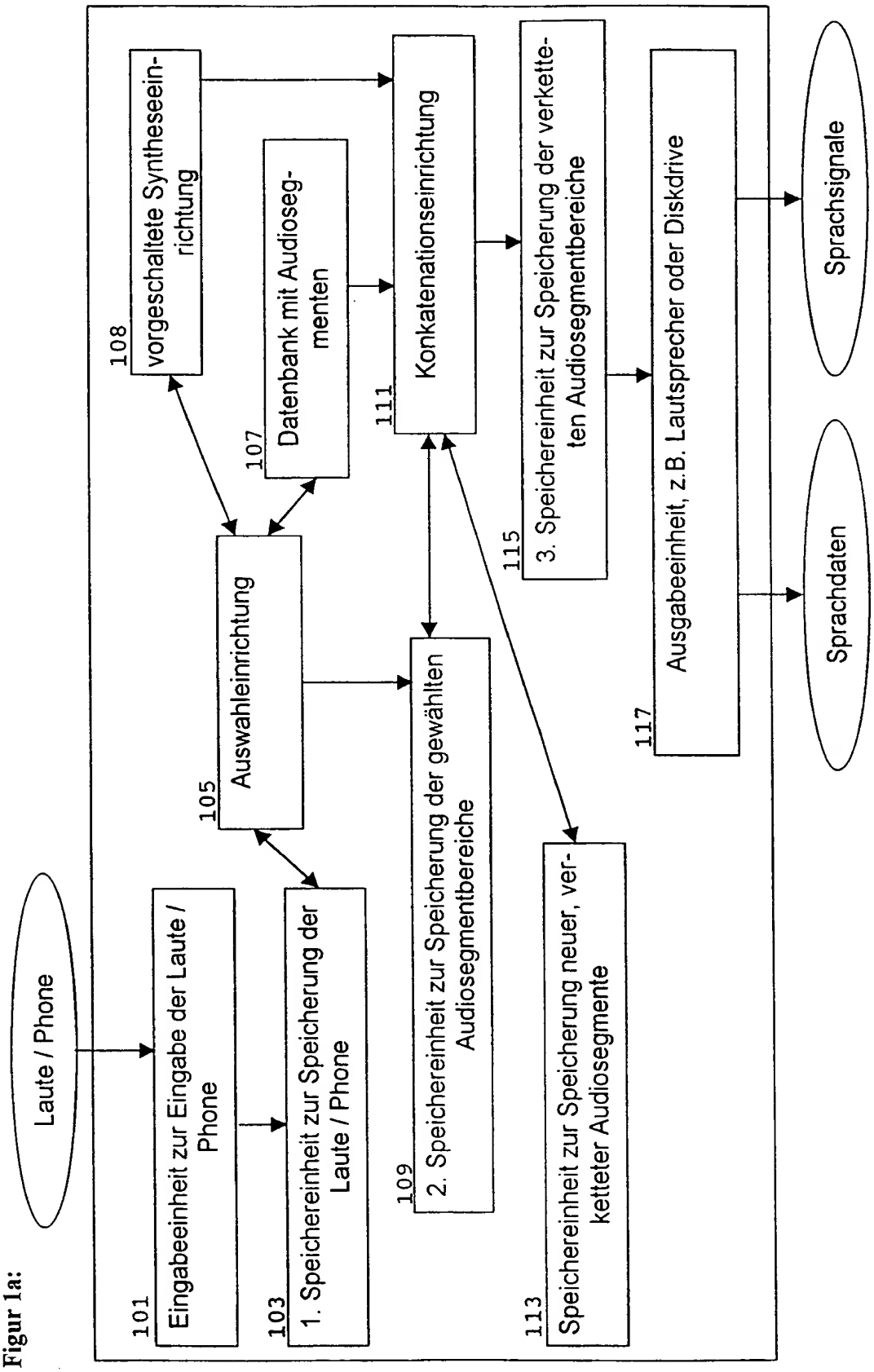
57. Datenspeicher nach einem der Ansprüche 50 bis 56, der zusätzlich Informationen enthält, die eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente und mit Hilfe geeigneter Funktionen in einem Bereich betreffen, in dem der Moment der Konkatenation liegt. Dies kann u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein.

58. Datenspeicher nach einem der Ansprüche 50 bis 57, der zusätzlich verkettete Audiosegmente zur Verfügung stellt, deren Moment der Konkatenation bei einer Stelle der verwendeten Bereiche des zeitlich vorgelagerten und/oder des zeitlich nachgelagerten Audiosegmentes liegt, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen. Diese Eigenschaften können u.a. sein: Nullstelle, Amplitudenwert, Steigung, Ableitung beliebigen Grades, Spektrum, Tonhöhe, Amplitudenwert in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften.

59. Datenspeicher nach einem der Ansprüche 50 bis 58, der zusätzlich Informationen in Form von heuristischem Wissen enthält, die die Auswahl der verwendeten Bereiche der einzelnen Audiosegmente, deren Bearbeitung, deren Variation sowie deren Konkatenation betreffen.

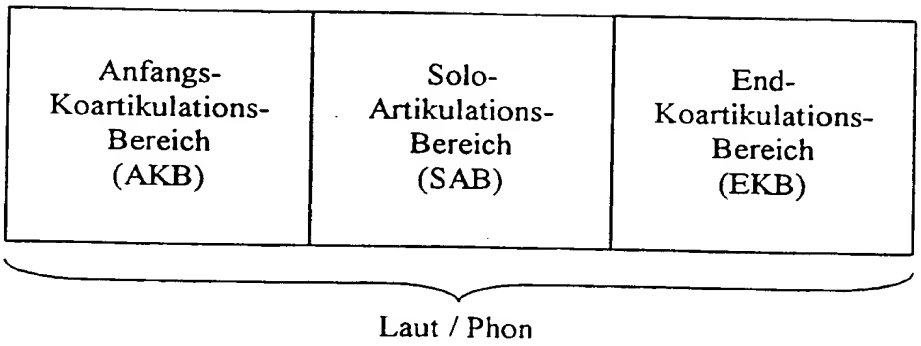
60. Tonträger, der Daten enthält, die zumindest teilweise synthetisierte akustische Daten sind, die
- mit einem Verfahren nach einem der Ansprüche 1 bis 13, oder
- mit einer Vorrichtung nach einem der Ansprüche 14 bis 27, oder
- unter Verwendung eines Datenträgers nach einem der Ansprüche 39 bis 49, oder
- unter Verwendung eines Datenspeichers nach einem der Ansprüche 50 bis 59
erzeugt wurden, oder
- die Sprachsignale nach einem der Ansprüche 28 bis 38 sind.

61. Tonträger nach Anspruch 60, dadurch gekennzeichnet, daß die synthetisierten akustischen Daten synthetisierte Sprachdaten sind.


Laute / Phone
101
Eingabeeinheit zur Eingabe der Laute / Phone
103
1. Speichereinheit zur Speicherung der Laute / Phone
105
Auswahleinrichtung
107
Datenbank mit Audiosegmenten
108
vorgeschaltete Syntheseeinrichtung
109
2. Speichereinheit zur Speicherung der gewählten Audiosegmentbereiche
111
Konkatenationseinrichtung
113
Speichereinheit zur Speicherung neuer, verketteter Audiosegmente
115
3. Speichereinheit zur Speicherung der verketteten Audiosegmentbereiche
117
Ausgabeeinheit, z.B. Lautsprecher oder Diskdrive
Sprachsignale
Sprachdaten

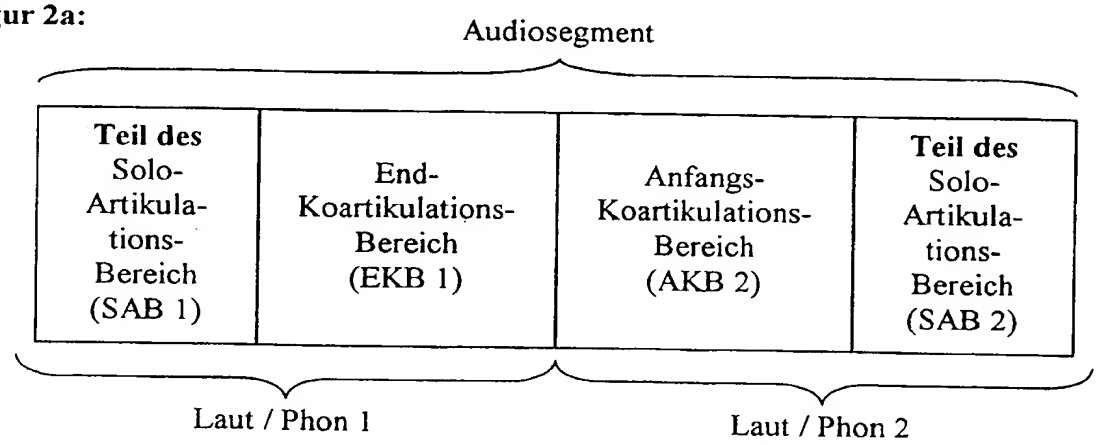
Figur 1a:

Figur 1b: Struktur eines **Lautes / Phons**.

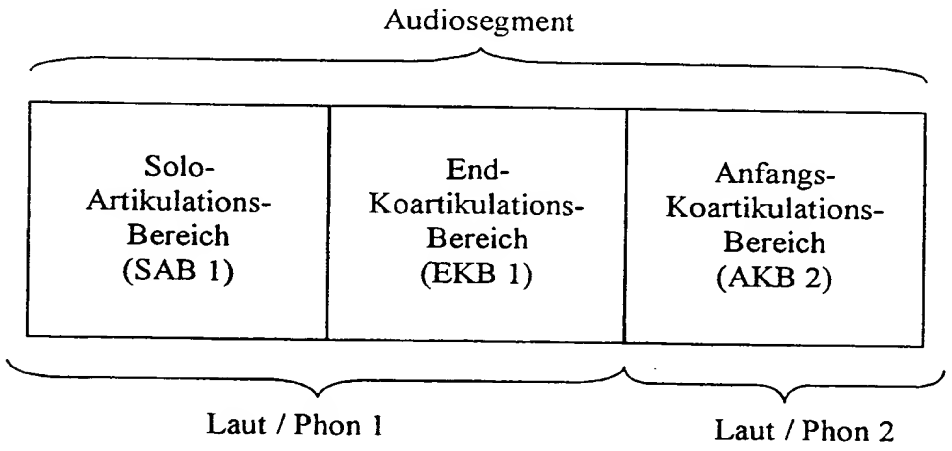


Figuren 2a bis 2i: Strukturen der Audiosegmente

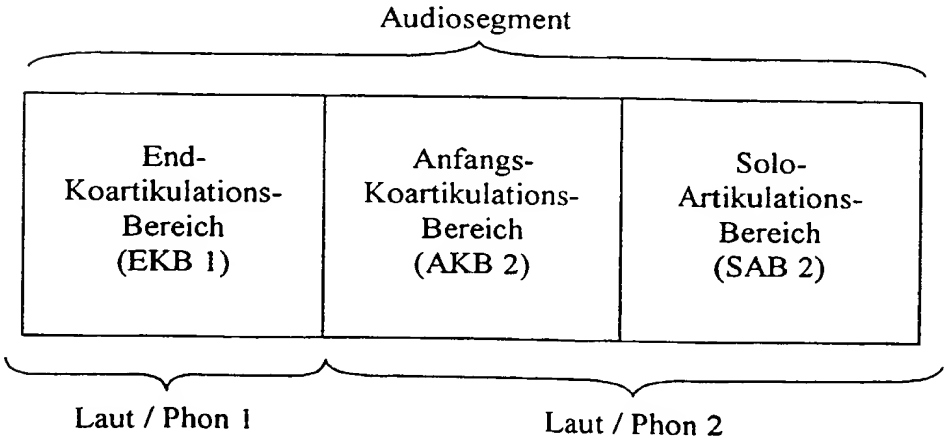
Figur 2a:



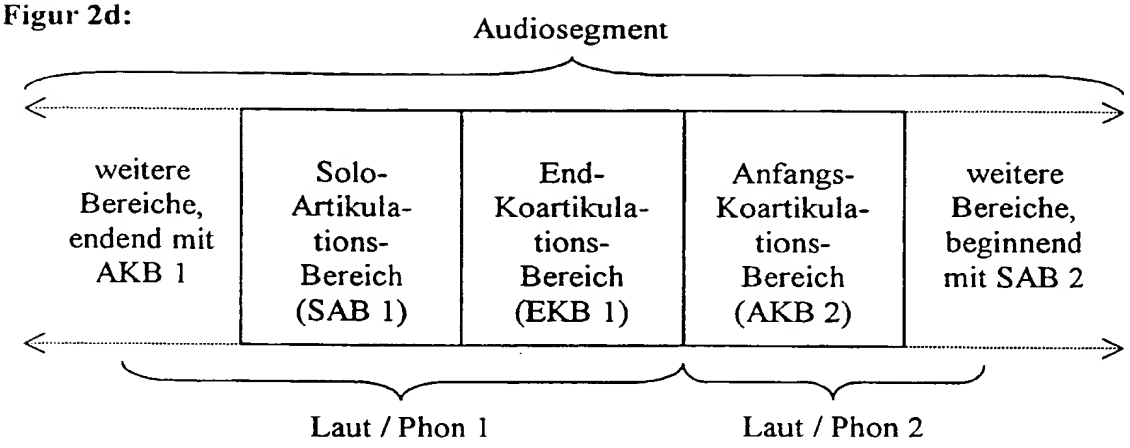
Figur 2b:



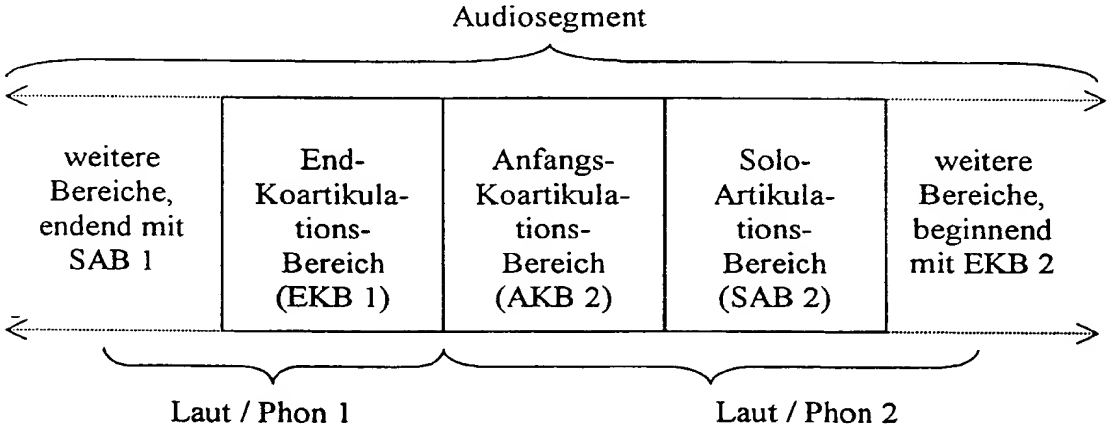
Figur 2c:



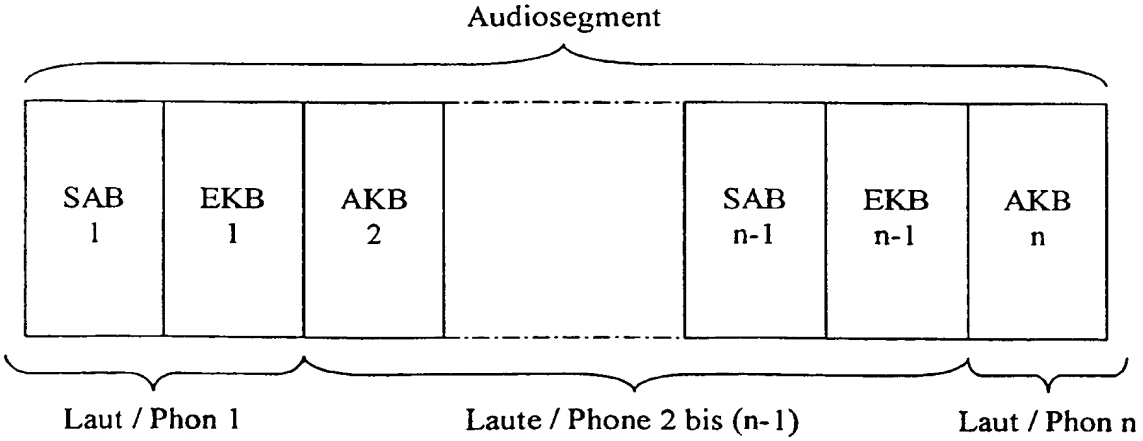
Figur 2d:



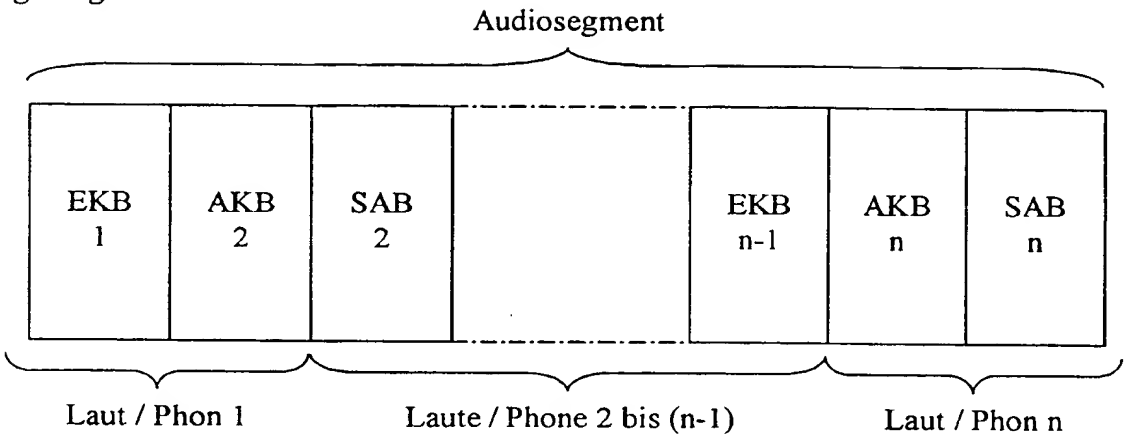
Figur 2e:



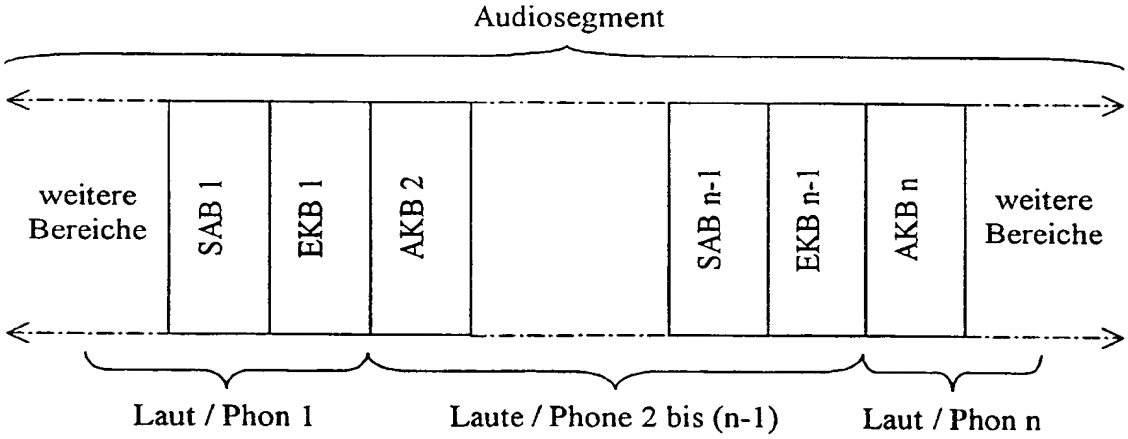
Figur 2f:



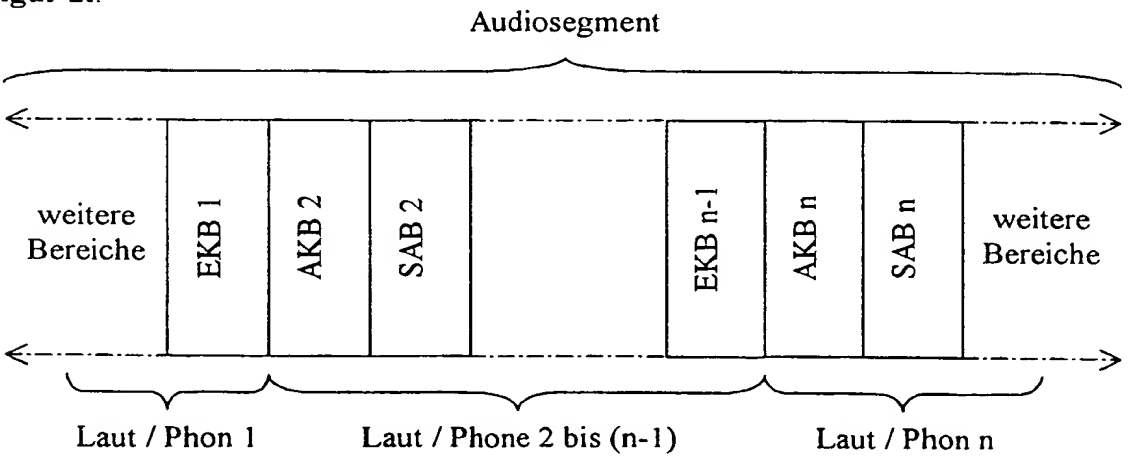
Figur 2g:



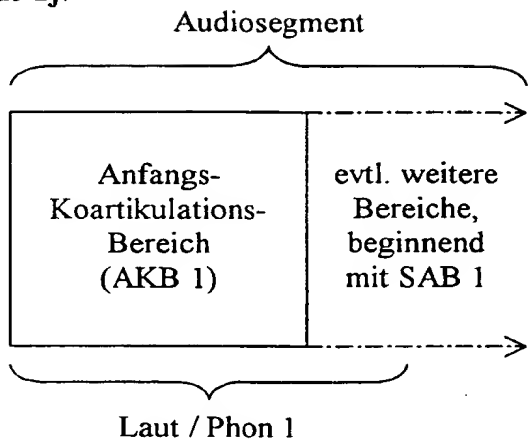
Figur 2h:



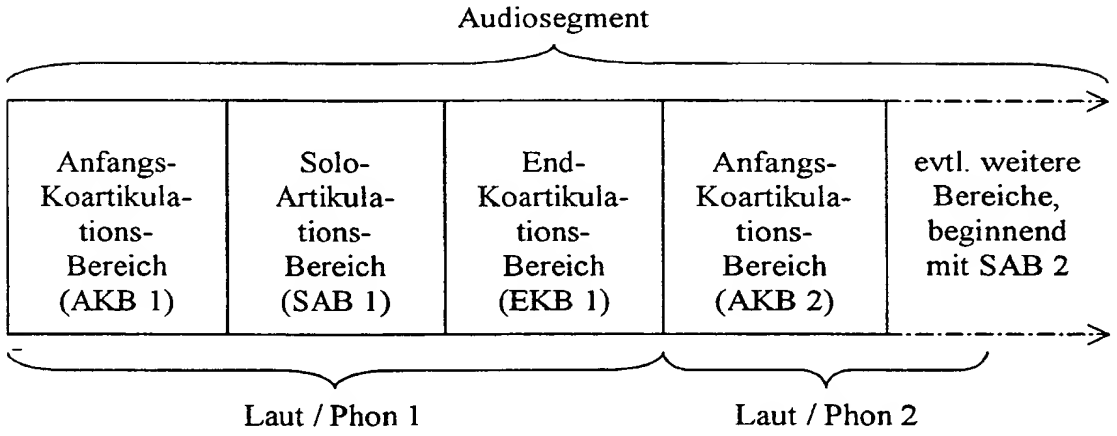
Figur 2i:



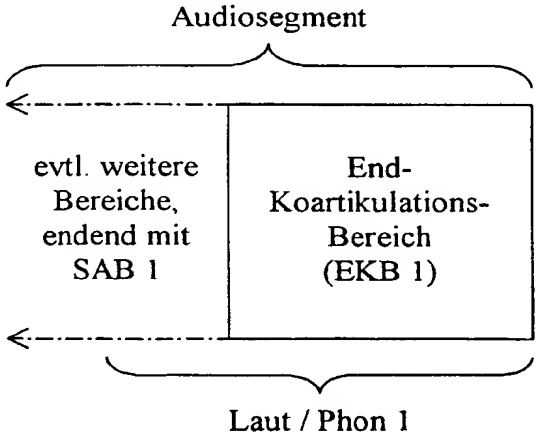
Figur 2j:



Figur 2k:

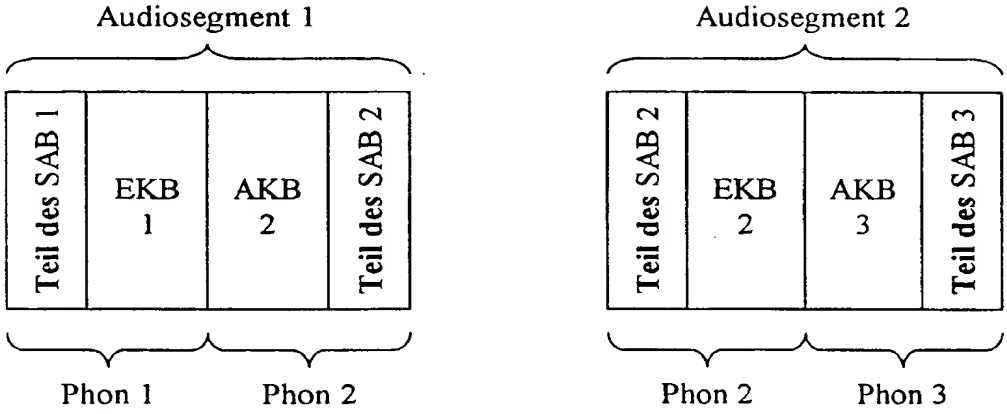


Figur 2l:

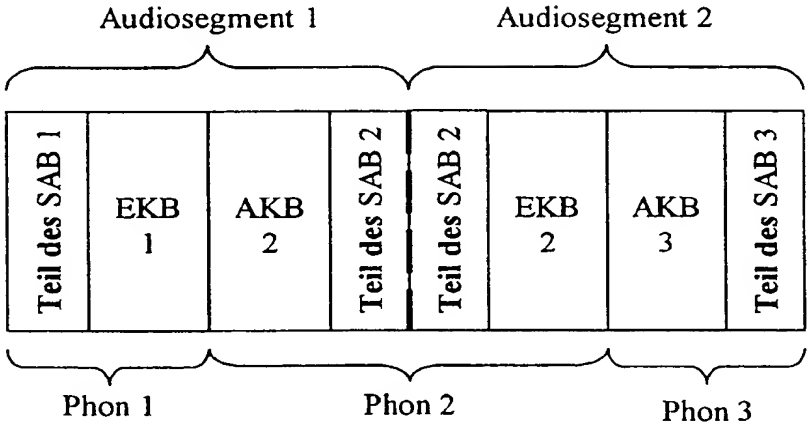


Figuren 3a bis 3d: Konkatenation

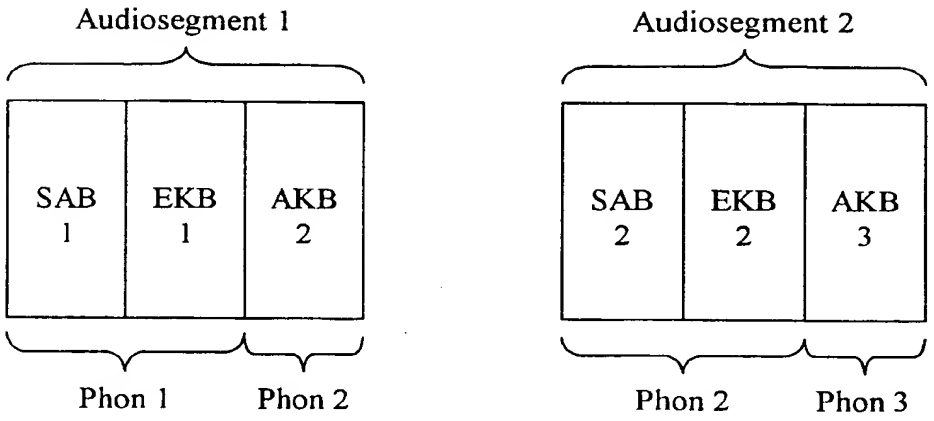
Figur 3a:



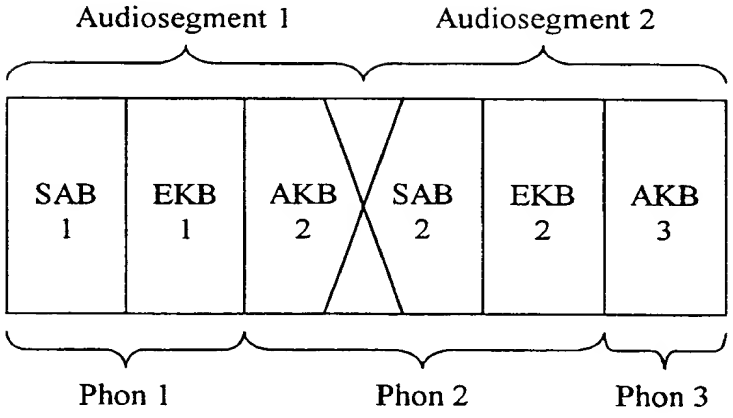
Figur 3aI:



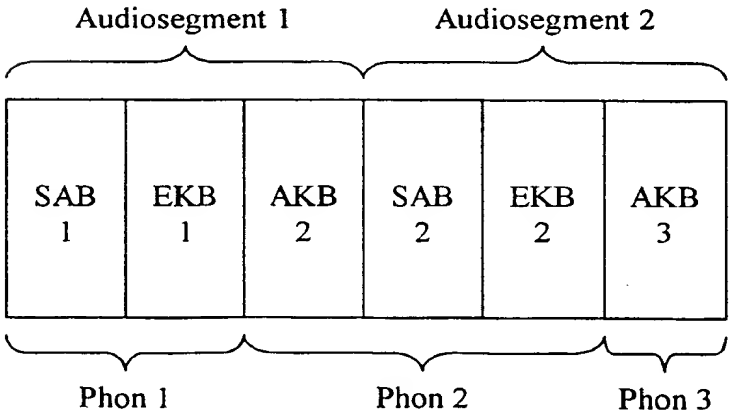
Figur 3b:



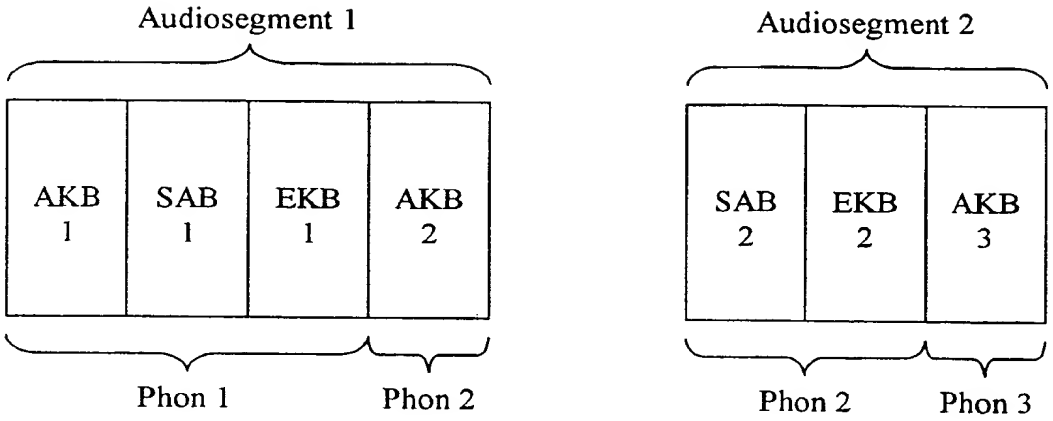
Figur 3bI:



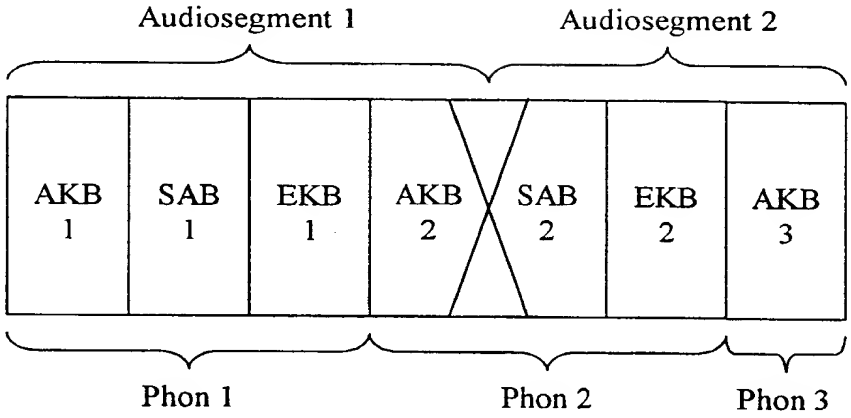
Figur 3bII:



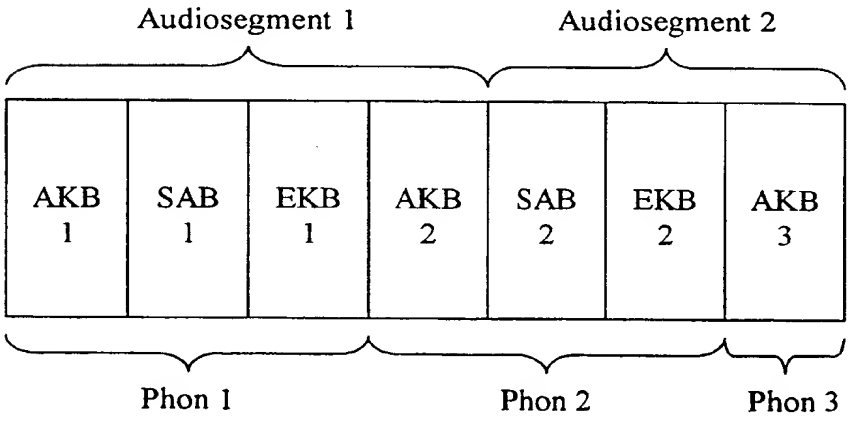
Figur 3c:



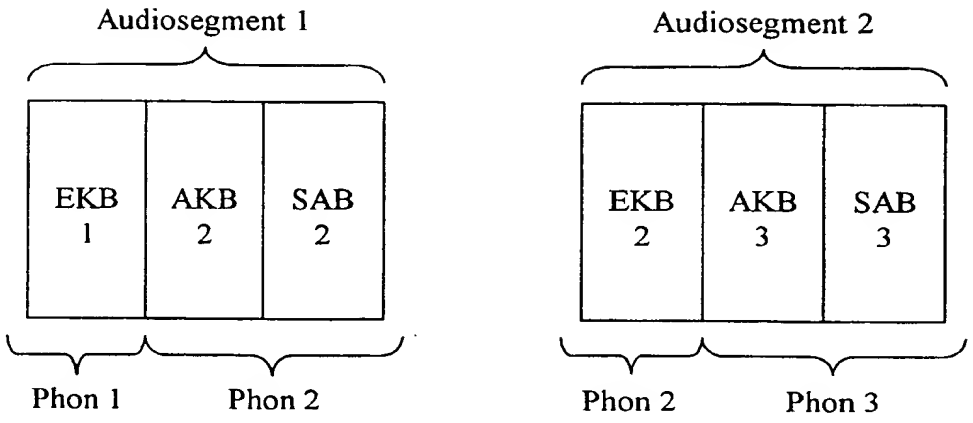
Figur 3cI:



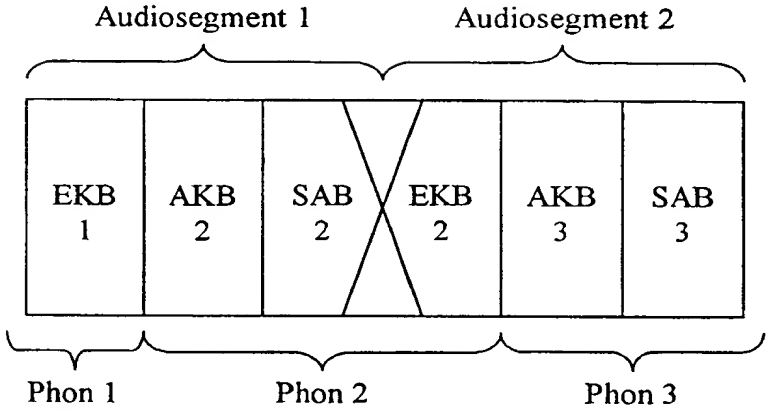
Figur 3cII:



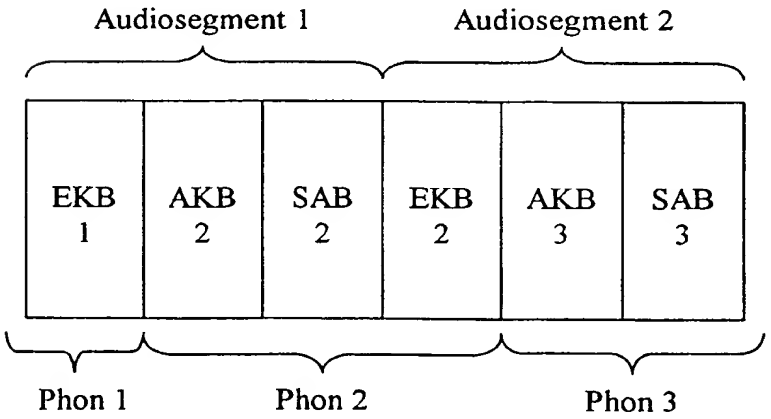
Figur 3d:



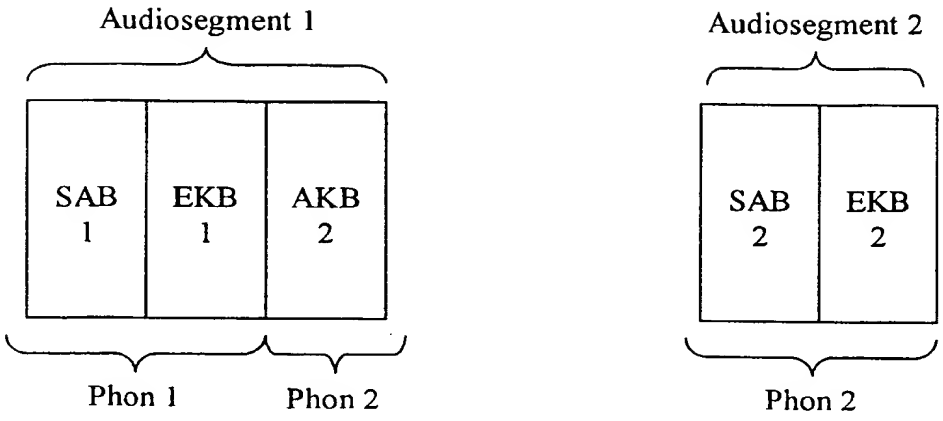
Figur 3dI:



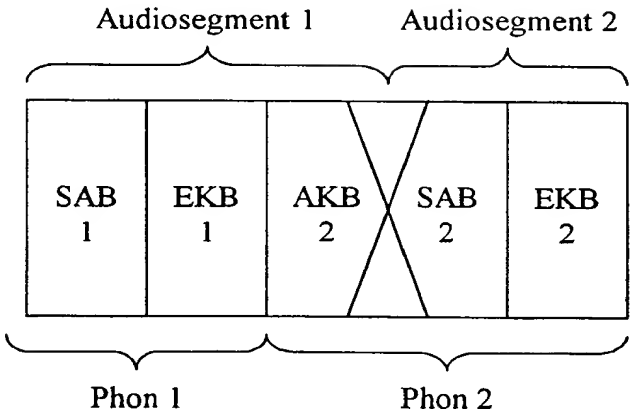
Figur 3dII:



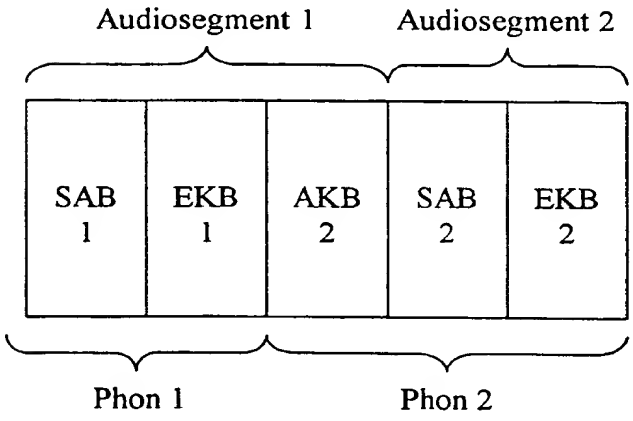
Figur 3e:



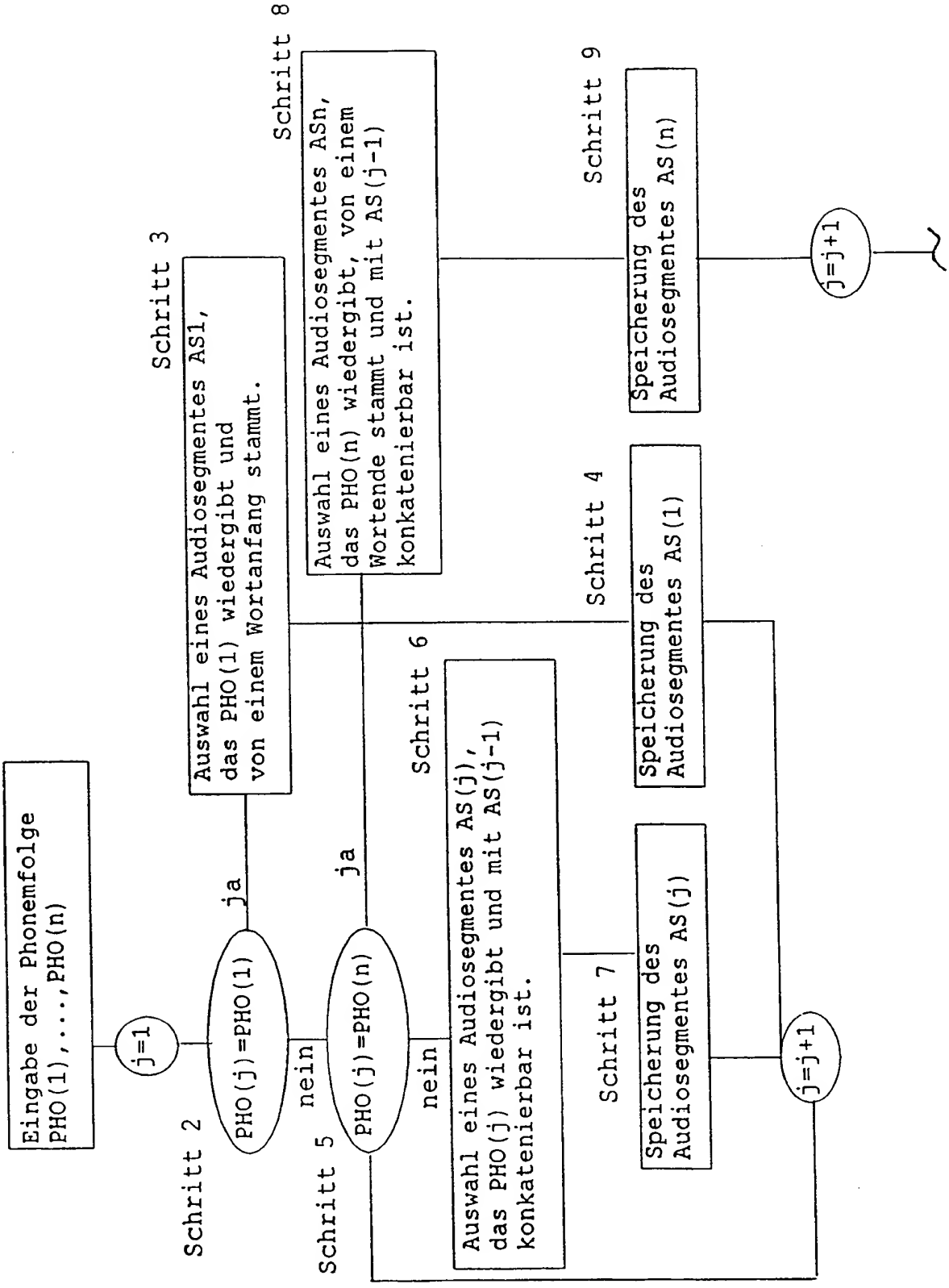
Figur 3eI:

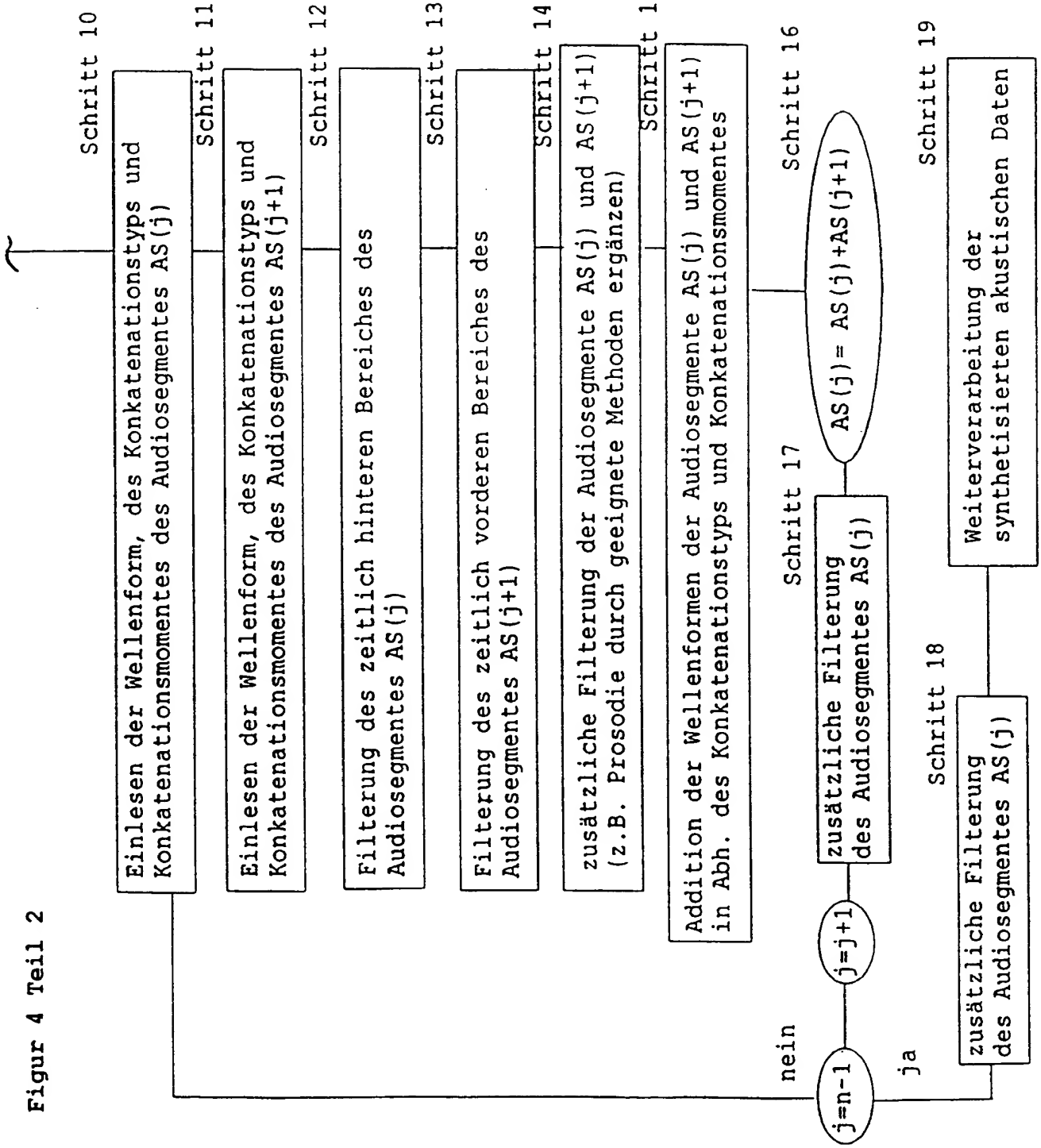


Figur 3eII:



Figur 4 Teil 1





Figur 4 Teil 2

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/EP 99/06081

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DETTWEILER H ET AL: "Concatenation rules for demisyllable speech synthesis" PROCEEDINGS OF IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP '85), TAMPA, FL, USA, vol. 2, 26 - 29 March 1985, pages 752-755, XP002128522 IEEE, New York, NY, USA the whole document	1-3, 14-16, 28-30, 39-41, 51-53, 55-61
A	US 5 659 664 A (KAJA JAAN) 19 August 1997 (1997-08-19) cited in the application column 3, line 44 -column 4, line 33	1-61
A	EP 0 351 848 A (SHARP KK) 24 January 1990 (1990-01-24) cited in the application abstract	1-61
A	EP 0 813 184 A (FACULTE POLYTECHNIQUE DE MONS) 17 December 1997 (1997-12-17) cited in the application column 8, line 19 -column 9, line 23	1-61
A	US 5 524 172 A (HAMON CHRISTIAN) 4 June 1996 (1996-06-04) cited in the application abstract; figures 3A,3B,3C,3D	1-61
A	WO 95 30193 A (MOTOROLA INC) 9 November 1995 (1995-11-09) cited in the application abstract	1-61

INTERNATIONAL SEARCH REPORT

Int tional Application No
PCT/EP 99/06081

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L13/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YIOURGALIS N ET AL: "A TtS system for the Greek language based on concatenation of formant coded segments" SPEECH COMMUNICATION,NL,ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, vol. 19, no. 1, page 21-38 XP004013506 ISSN: 0167-6393 page 22 -page 32 --- -/--	1-3, 14-16, 28-30, 39-41, 51-53, 55-61

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 January 2000	04/02/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Ramos Sánchez, U

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte onal Application No
PCT/EP 99/06081

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5659664 A	19-08-1997	SE 469576 B	26-07-1993
		DE 69318209 D	04-06-1998
		DE 69318209 T	27-08-1998
		EP 0561752 A	22-09-1993
		GB 2265287 A,B	22-09-1993
		JP 6041557 A	15-02-1994
		SE 9200817 A	26-07-1993
EP 0351848 A	24-01-1990	JP 1999183 C	08-12-1995
		JP 2032399 A	02-02-1990
		JP 7027397 B	29-03-1995
		DE 68915353 D	23-06-1994
		DE 68915353 T	20-10-1994
		US 5111505 A	05-05-1992
EP 0813184 A	17-12-1997	BE 1010336 A	02-06-1998
US 5524172 A	04-06-1996	FR 2636163 A	09-03-1990
		CA 1324670 A	23-11-1993
		DE 68919637 D	12-01-1995
		DE 68919637 T	20-07-1995
		DK 107390 A	30-05-1990
		EP 0363233 A	11-04-1990
		ES 2065406 T	16-02-1995
		WO 9003027 A	22-03-1990
		JP 3501896 T	25-04-1991
		US 5327498 A	05-07-1994
WO 9530193 A	09-11-1995	AU 675389 B	30-01-1997
		AU 2104095 A	29-11-1995
		CA 2161540 A	09-11-1995
		CN 1128072 A	31-07-1996
		EP 0710378 A	08-05-1996
		FI 955608 A	22-11-1995
		JP 8512150 T	17-12-1996
		US 5668926 A	16-09-1997

INTERNATIONALER RECHERCHENBERICHT

Int. .tionales Aktenzeichen
PCT/EP 99/06081

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
IPK 7 G10L13/06

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
IPK 7 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	YIOURGALIS N ET AL: "A TtS system for the Greek language based on concatenation of formant coded segments" SPEECH COMMUNICATION,NL,ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, Bd. 19, Nr. 1, Seite 21-38 XP004013506 ISSN: 0167-6393 Seite 22 -Seite 32 --- -/--	1-3, 14-16, 28-30, 39-41, 51-53, 55-61

[X] Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

[X] Siehe Anhang Patentfamilie

° Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absendedatum des internationalen Recherchenberichts
24. Januar 2000	04/02/2000

Name und Postanschrift der Internationalen Recherchenbehörde	Bevollmächtigter Bediensteter
Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Ramos Sánchez, U

INTERNATIONALER RECHERCHENBERICHT

Int. tionales Aktenzeichen
PCT/EP 99/06081

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	DETTWEILER H ET AL: "Concatenation rules for demisyllable speech synthesis" PROCEEDINGS OF IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP '85), TAMPA, FL, USA, Bd. 2, 26. - 29. März 1985, Seiten 752-755, XP002128522 IEEE, New York, NY, USA das ganze Dokument ---	1-3, 14-16, 28-30, 39-41, 51-53, 55-61
A	US 5 659 664 A (KAJA JAAN) 19. August 1997 (1997-08-19) in der Anmeldung erwähnt Spalte 3, Zeile 44 -Spalte 4, Zeile 33 ---	1-61
A	EP 0 351 848 A (SHARP KK) 24. Januar 1990 (1990-01-24) in der Anmeldung erwähnt Zusammenfassung ---	1-61
A	EP 0 813 184 A (FACULTE POLYTECHNIQUE DE MONS) 17. Dezember 1997 (1997-12-17) in der Anmeldung erwähnt Spalte 8, Zeile 19 -Spalte 9, Zeile 23 ---	1-61
A	US 5 524 172 A (HAMON CHRISTIAN) 4. Juni 1996 (1996-06-04) in der Anmeldung erwähnt Zusammenfassung; Abbildungen 3A,3B,3C,3D ---	1-61
A	WO 95 30193 A (MOTOROLA INC) 9. November 1995 (1995-11-09) in der Anmeldung erwähnt Zusammenfassung -----	1-61

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen
PCT/EP 99/06081

im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 5659664 A	19-08-1997	SE 469576 B	26-07-1993
		DE 69318209 D	04-06-1998
		DE 69318209 T	27-08-1998
		EP 0561752 A	22-09-1993
		GB 2265287 A,B	22-09-1993
		JP 6041557 A	15-02-1994
		SE 9200817 A	26-07-1993
EP 0351848 A	24-01-1990	JP 1999183 C	08-12-1995
		JP 2032399 A	02-02-1990
		JP 7027397 B	29-03-1995
		DE 68915353 D	23-06-1994
		DE 68915353 T	20-10-1994
		US 5111505 A	05-05-1992
EP 0813184 A	17-12-1997	BE 1010336 A	02-06-1998
US 5524172 A	04-06-1996	FR 2636163 A	09-03-1990
		CA 1324670 A	23-11-1993
		DE 68919637 D	12-01-1995
		DE 68919637 T	20-07-1995
		DK 107390 A	30-05-1990
		EP 0363233 A	11-04-1990
		ES 2065406 T	16-02-1995
		WO 9003027 A	22-03-1990
		JP 3501896 T	25-04-1991
		US 5327498 A	05-07-1994
WO 9530193 A	09-11-1995	AU 675389 B	30-01-1997
		AU 2104095 A	29-11-1995
		CA 2161540 A	09-11-1995
		CN 1128072 A	31-07-1996
		EP 0710378 A	08-05-1996
		FI 955608 A	22-11-1995
		JP 8512150 T	17-12-1996
		US 5668926 A	16-09-1997

09/763149

VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS

PCT

INTERNATIONALER VORLÄUFIGER PRÜFUNGSBERICHT

(Artikel 36 und Regel 70 PCT)

Aktenzeichen des Anmelders oder Anwalts	WEITERES VORGEHEN	siehe Mitteilung über die Übersendung des internationalen vorläufigen Prüfungsbericht (Formblatt PCT/IPEA/416)
EP-82 972/PC		
Internationales Aktenzeichen	Internationales Anmeldedatum *(Tag/Monat/Jahr)*	Prioritätsdatum *(Tag/Monat/Tag)*
PCT/EP99/06081	19/08/1999	19/08/1998
Internationale Patentklassifikation (IPK) oder nationale Klassifikation und IPK		
G10L13/06		
Anmelder		
BUSKIES, Christoph		

1. Dieser internationale vorläufige Prüfungsbericht wurde von der mit der internationale vorläufigen Prüfung beauftragte Behörde erstellt und wird dem Anmelder gemäß Artikel 36 übermittelt.

2. Dieser BERICHT umfaßt insgesamt 6 Blätter einschließlich dieses Deckblatts.

   ☒ Außerdem liegen dem Bericht ANLAGEN bei; dabei handelt es sich um Blätter mit Beschreibungen, Ansprüchen und/oder Zeichnungen, die geändert wurden und diesem Bericht zugrunde liegen, und/oder Blätter mit vor dieser Behörde vorgenommenen Berichtigungen (siehe Regel 70.16 und Abschnitt 607 der Verwaltungsrichtlinien zum PCT).

   Diese Anlagen umfassen insgesamt 14 Blätter.

3. Dieser Bericht enthält Angaben zu folgenden Punkten:

   - I ☒ Grundlage des Berichts
   - II ☐ Priorität
   - III ☒ Keine Erstellung eines Gutachtens über Neuheit, erfinderische Tätigkeit und gewerbliche Anwendbarkeit
   - IV ☐ Mangelnde Einheitlichkeit der Erfindung
   - V ☒ Begründete Feststellung nach Artikel 35(2) hinsichtlich der Neuheit, der erfinderische Tätigkeit und der gewerbliche Anwendbarkeit; Unterlagen und Erklärungen zur Stützung dieser Feststellung
   - VI ☐ Bestimmte angeführte Unterlagen
   - VII ☒ Bestimmte Mängel der internationalen Anmeldung
   - VIII ☒ Bestimmte Bemerkungen zur internationalen Anmeldung

Datum der Einreichung des Antrags	Datum der Fertigstellung dieses Berichts
17/03/2000	02.10.2000
Name und Postanschrift der mit der internationalen vorläufigen Prüfung beauftragten Behörde:	Bevollmächtigter Bediensteter
Europäisches Patentamt D-80298 München Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	De Vos, L Tel. Nr. +49 89 2399 2048

Formblatt PCT/IPEA/409 (Deckblatt) (Januar 1994)

INTERNATIONALER VORLÄUFIGER PRÜFUNGSBERICHT

Internationales Aktenzeichen PCT/EP99/06081

I. Grundlage des Berichts

1. Dieser Bericht wurde erstellt auf der Grundlage (*Ersatzblätter, die dem Anmeldeamt auf eine Aufforderung nach Artikel 14 hin vorgelegt wurden, gelten im Rahmen dieses Berichts als "ursprünglich eingereicht" und sind ihm nicht beigefügt, weil sie keine Änderungen enthalten.)*:

Beschreibung, Seiten:

1-22 ursprüngliche Fassung

Patentansprüche, Nr.:

1-68 eingegangen am 24/08/2000 mit Schreiben vom 24/08/2000

Zeichnungen, Blätter:

1/13-13/13 ursprüngliche Fassung

2. Aufgrund der Änderungen sind folgende Unterlagen fortgefallen:

☐ Beschreibung, Seiten:
☐ Ansprüche, Nr.:
☐ Zeichnungen, Blatt:

3. ☐ Dieser Bericht ist ohne Berücksichtigung (von einigen) der Änderungen erstellt worden, da diese aus den angegebenen Gründen nach Auffassung der Behörde über den Offenbarungsgehalt in der ursprünglich eingereichten Fassung hinausgehen (Regel 70.2(c)):

4. Etwaige zusätzliche Bemerkungen:

III. Keine Erstellung eines Gutachtens über Neuheit, erfinderische Tätigkeit und gewerbliche Anwendbarkeit

Folgende Teile der Anmeldung wurden nicht daraufhin geprüft, ob die beanspruchte Erfindung als neu, auf erfinderischer Tätigkeit beruhend (nicht offensichtlich) und gewerblich anwendbar anzusehen ist:

☐ die gesamte internationale Anmeldung.

☒ Ansprüche Nr. 58-66.

Begründung:

☐ Die gesamte internationale Anmeldung, bzw. die obengenannten Ansprüche Nr. beziehen sich auf den nachstehenden Gegenstand, für den keine internationale vorläufige Prüfung durchgeführt werden braucht (*genaue Angaben*):

☒ Die Beschreibung, die Ansprüche oder die Zeichnungen (*machen Sie hierzu nachstehend genaue Angaben*) oder die obengenannten Ansprüche Nr. 58-66 sind so unklar, daß kein sinnvolles Gutachten erstellt werden konnte (*genaue Angaben*):

siehe Beiblatt

☐ Die Ansprüche bzw. die obengenannten Ansprüche Nr. sind so unzureichend durch die Beschreibung gestützt, daß kein sinnvolles Gutachten erstellt werden konnte.

☐ Für die obengenannten Ansprüche Nr. wurde kein internationaler Recherchenbericht erstellt.

V. Begründete Feststellung nach Artikel 35(2) hinsichtlich der Neuheit, der erfinderischen Tätigkeit und der gewerblichen Anwendbarkeit; Unterlagen und Erklärungen zur Stützung dieser Feststellung

1. Feststellung

Neuheit (N)	Ja:	Ansprüche	1-57, 67-68
	Nein:	Ansprüche	
Erfinderische Tätigkeit (ET)	Ja:	Ansprüche	1-57, 67-68
	Nein:	Ansprüche	
Gewerbliche Anwendbarkeit (GA)	Ja:	Ansprüche	1-57, 67-68
	Nein:	Ansprüche	

2. Unterlagen und Erklärungen

siehe Beiblatt

VII. Bestimmte Mängel der internationalen Anmeldung

Es wurde festgestellt, daß die internationale Anmeldung nach Form oder Inhalt folgende Mängel aufweist:

siehe Beiblatt

VIII. Bestimmte Bemerkungen zur internationalen Anmeldung

Zur Klarheit der Patentansprüche, der Beschreibung und der Zeichnungen oder zu der Frage, ob die Ansprüche in vollem Umfang durch die Beschreibung gestützt werden, ist folgendes zu bemerken:

siehe Beiblatt

III. Keine Erstellung eines Gutachtens

1. Der unabhängige Patentanspruch 58 und die von diesem Anspruch abhängigen Ansprüche 59-66 beanspruchen einen Datenspeicher, welcher Audiosegmente enthält. Diese Audiosegmente werden jedoch nur durch das zu erreichende Ergebnis gekennzeichnet, nämlich, daß sie geeignet sein müssen, synthetisierte akustische Daten unter Verwendung des Verfahrens nach Anspruch 1, der Vorrichtung nach Anspruch 16 oder des Datenträgers nach Anspruch 33 zu erzeugen.

   Da hierdurch keine Merkmale der Audiosegmente im Anspruch festgelegt werden (R. 6.3.a PCT), ist der Anspruch nicht klar (Art. 6 PCT), s. auch die Richtlinien für die PCT-Prüfung, III-4.7.

2. Folglich werden die Ansprüche 58-66 gemäß Art. 34.4(a)(ii) PCT von der Erstellung eines Gutachtens ausgenommen.

V. Begründete Feststellung nach Art. 35(2) PCT

3. Die vorliegende Anmeldung befaßt sich mit Sprachsynthese. Insbesondere befaßt sie sich mit datenbasierten Systemen, bei denen die saubere Konkatenation der einzelnen im Speicher vorhandenen Sprachmusterstücke sich als sehr kritisch für die erreichte Synthesequalität erweist.

   Der nächste Stand der Technik ist im Dokument "A TtS system for the Greek language based on the concatenation of formant coded segments", Yourgalis et. al., Speech Communication 19(1996), S. 21-38 offenbart. Dieses Dokument zeigt, daß durch entsprechende Simulationsberechnungen der Koartikulationseffekt berücksichtigt werden kann, und daß je nach Art der Phoneme andere Konkatenationsweisen verwendet werden können, die jeweils in ihrer Klasse bessere Ergebnisse liefern können als die anderen zur Auswahl stehenden Methoden.

   Das technische Problem, das in der vorliegenden Anmeldung zu lösen ist, ist das

Finden einer Alternative zu den im nächsten Stand der Technik vorhandenen Lösungsansätzen.

Hierzu wird die Koartikulationsproblematik vollständig datenbasiert angegangen. Der zu verwendende Bereich des zeitlich nachgelagerten Audiosegments (an das bisherige Ergebnis hinten anzuhängen) endet hierbei mit dem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich dieses zeitlich nachgelagerten Audiosegments.

Weiter wird nicht die Weise der Konkatenation, sondern der Moment (Zeitpunkt) der Konkatenation in der vorliegenden Anmeldung in Abhängigkeit der Eigenschaften angrenzenden Bereiche situationsabhängig bestimmt.

Dies wird beansprucht in unabhängigen Verfahrens-, Vorrichtungs- und Programmdatenträger-Ansprüchen 1, 16 und 33.

Die Verwendung der in Anspruch 1 aufgelisteten Schritte wird somit in Dokument D1 weder veröffentlicht noch nahegelegt. Keines der übrigen Dokumente des Internationalen Recherchenberichtes enthält eine Andeutung, die den Fachmann dazu veranlassen würde, den Stand der Technik in Dokument D1 durch die Merkmale des Anspruchs 1 zu ersetzen.

Anspruch 1 ist somit neu und beruht auf einer erfinderischen Tätigkeit im Sinne des Art. 33(2)-(3) PCT.

Dies gilt ebenfalls für den unabhängigen Vorrichtungsanspruch 16 und den unabhängigen Programmdatenträgeranspruch 33. Gleiches gilt für den unabhängigen Erzeugnisanspruch 47 und für den unabhängigen Erzeugnisdatenträgeranspruch 67, sofern dieser sich nicht auf den unklaren Anspruch 58 bezieht. Desweiteren gilt dies auch für die sich auf diese unabhängigen Ansprüche beziehenden abhängigen Ansprüche 2-15, 17-32, 34-46, 48-57 und 68 (sofern Anspruch 67 sich nicht auf Anspruch 58 bezieht), da diese keinen Widerspruch zu den unabhängigen Ansprüchen verursachen, auf welche sie sich beziehen.

VII. Bestimmte Mängel der Internationalen Anmeldung.

4. Im Widerspruch zu den Erfordernissen der Regel 5.1 a) ii) PCT werden in der Beschreibung weder der in dem Dokument D1 offenbarte einschlägige Stand der Technik noch dieses Dokument angegeben.

VIII. Bestimmte Bemerkungen zur Internationalen Anmeldung.

5. In der Beschreibung, S. 9, im Satz von Z. 32 bis Z. 37, fehlt mindestens ein Wort. Der Satz ist dadurch nicht verständlich. Der Prüfer nimmt an, daß "bestimmt" auf Z. 37 durch "bestimmt wird" hätte ersetzt werden sollen.

6. Zwischen S. 10, Z. 21-30 und S. 12, Z. 14-20 einerseits, und S. 17, Z. 24-30 und S. 18, Z. 25-34 andererseits, herrscht eine gewisse Unstimmigkeit. In den erstgenannten Passagen werden Audiosegmentbereiche bevorzugt, welche mit der Wiedergabe eines dynamischen Lauts beginnen, während in den letztgenannten Textabschnitten genau gegenteilig das Anfangen mit einem statischen Laut bevorzugt wird. Dem Leser bleibt in diesem Fall die Frage, in welchem der beiden Fällen der (in beiden Fällen) angedeutete Vorteil eines geringeren Aufwandes auch tatsächlich vorhanden ist. Diese Unklarheit hätte beseitigt werden müssen.

7. Die verwendeten Nummern der Ansprüche in der Beschreibung stimmen nicht mit der tatsächlichen Numerierung überein.

Ansprüche

1. Verfahren zur koartikulationsgerechten Konkatenation von Audiosegmenten, um synthetisierte akustische Daten zu erzeugen, die eine Folge konkatenierter Laute wiedergeben, mit folgenden Schritten:
- Auswahl von wenigstens zwei Audiosegmenten, die Bereiche enthalten, die jeweils einen Teil eines Lautes oder einen Teil der Lautfolge wiedergeben, aufweist,
gekennzeichnet durch die Schritte:
- Festlegen eines zu verwendenden Bereiches eines zeitlich vorgelagerten Audiosegments,
- Festlegen eines zu verwendenden Bereiches eines zeitlich nachgelagerten Audiosegments, der zeitlich mit dem zeitlich nachgelagerten Audiosegment beginnt und mit dem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich des zeitlich nachgelagerten Audiosegments endet,
- wobei die Dauer und Lage der zu verwendenden Bereiche in Abhängigkeit der vor- und nachgelagerten Audiosegmente bestimmt wird, und
- Konkatenieren des festgelegten Bereiches des zeitlich vorgelagerten Audiosegments mit dem festgelegten Bereich des zeitlich nachgelagerten Audiosegments, indem der Moment der Konkatenation in Abhängigkeit von Eigenschaften des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments in einen Bereich gelegt wird, der zeitlich unmittelbar vor dem verwendeten Bereich des zeitlich nachgelagerten Audiosegments beginnt und mit diesem endet.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß
- der Moment der Konkatenation in einen Bereich gelegt wird, der in der Umgebung der Grenzen des zuerst zu verwendenden Soloartikulationsbereichs des zeitlich nachgelagerten Audiosegments liegt, wenn dessen zu verwendender Bereich am Anfang einen statischen Laut wiedergibt, und
- ein zeitlich hinterer Bereich des zu verwendenden Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des zu verwendenden Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und überlappend addiert werden (Crossfade), wobei die Übergangsfunktionen und die Länge eines Überlappungsbereichs der beiden Bereiche in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt werden.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß
- der Moment der Konkatenation in einen Bereich gelegt wird, der zeitlich unmittelbar vor dem zu verwendenden Bereich des zeitlich nachgelagerten Audiosegments liegt, wenn dessen verwendeter Bereich am Anfang einen dynamischen Laut wiedergibt, und
- ein zeitlich hinterer Bereich des zu verwendenden Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des zu verwendenden Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und nicht überlappend verbunden werden (Hardfade), wobei die Übergangsfunktionen in Abhängigkeit der zu synthetisierenden akustischen Daten bestimmt werden.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß für einen Laut oder einen Teil der Folge konkatenierter Laute am Anfang der konkatenierten Lautfolge ein Bereich eines Audiosegmentes ausgewählt wird, so daß der Anfang des Bereiches die Eigenschaften des Anfangs der konkatenierten Lautfolge wiedergibt.

5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß für einen Laut oder einen Teil der Folge konkatenierter Laute am Ende der konkatenierten Lautfolge ein Bereich eines Audiosegmentes ausgewählt wird, so daß das Ende des Bereiches die Eigenschaften des Endes der konkatenierten Lautfolge wiedergibt.

6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß die zu synthetisierenden Sprachdaten in Gruppen zusammengefaßt werden, die jeweils durch ein einzelnes Audiosegment beschrieben werden.

7. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, daß für den zeitlich nachgelagerten Audiosegmentbereich ein Audiosegmentbereich gewählt wird, der die größte Anzahl aufeinanderfolgender Teile der Laute der Lautfolge wiedergibt, um bei der Erzeugung der synthetisierten akustischen Daten die kleinste Anzahl von Audiosegmentbereichen zu verwenden.

8. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der konkatenierten Lautfolge durchgeführt wird, wobei die Eigenschaften u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums sein können.

9. Verfahren nach einem der Ansprüche 1 bis 8, dadurch gekennzeichnet, daß eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in einem Bereich durchgeführt wird, in dem der Moment der Konkatenation liegt, wobei die Funktionen u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums betreffen können.

10. Verfahren nach einem der Ansprüche 1 bis 9, dadurch gekennzeichnet, daß der Moment der Konkatenation an Stellen in den zu verwendenden Bereichen des zeitlich vorgelagerten und/oder des zeitlich nachgelagerten Audiosegments gelegt wird, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen, wobei die Eigenschaften u.a. Nullstellen, Amplitudenwerte, Steigungen, Ableitungen beliebigen Grades, Spektren, Tonhöhen, Amplitudenwerte in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften sein können.

11. Verfahren nach einem der Ansprüche 1 bis 10, dadurch gekennzeichnet, daß
- die Auswahl der verwendeten Bereiche einzelner Audiosegmente, deren Bearbeitung, deren Variation sowie deren Konkatenation zusätzlich unter Verwendung heuristischen Wissens durchgeführt wird, das durch ein zusätzlich durchgeführtes heuristisches Verfahren gewonnen wird.

12. Verfahren einem der Ansprüche 1 bis 11, dadurch gekennzeichnet, daß
- die zu synthetisierenden akustischen Daten Sprachdaten und die Laute Phone sind.

13. Verfahren nach einem der Ansprüche 2 bis 12, dadurch gekennzeichnet, daß
- die statischen Laute Vokale, Diphtonge, Liquide, Vibranten, Frikative und Nasale umfassen.

14. Verfahren nach einem der Ansprüche 3 bis 13, dadurch gekennzeichnet, daß
- die dynamischen Laute Plosive, Affrikate, Glottalstops und geschlagenen Laute umfassen.

15. Verfahren nach einem der Ansprüche 1 bis 14, dadurch gekennzeichnet, daß
- eine Umwandlung der synthetisierten akustischen Daten in akustische Signale und/oder Sprachsignale durchgeführt wird.

16. Vorrichtung zur koartikulationsgerechten Konkatenation von Audiosegmenten, um synthetisierte akustische Daten zu erzeugen, die eine Folge von Lauten wiedergeben, mit:
- einer Datenbank (107), in der Audiosegmente gespeichert sind, die jeweils Teile eines Lautes oder Teile einer Folge von (konkatenierten) Lauten wiedergeben
- und/oder einer beliebigen vorgeschalteten Syntheseeinrichtung (108), die Audiosegmente liefert,
- einer Einrichtung (105) zur Auswahl von wenigstens zwei Audiosegmenten aus der Datenbank (107) und/oder der vorgeschalteten Syntheseeinrichtung (108), und
- einer Einrichtung (111) zur Konkatenation der Audiosegmente, dadurch gekennzeichnet, daß die Konkatenationseinrichtung (111) geeignet ist,
- einen zu verwendenden Bereiches eines zeitlich vorgelagerten Audiosegments zu definieren,
- einen zu verwendenden Bereiches eines zeitlich nachgelagerten Audiosegments in einem Bereich zu definieren, der mit dem zeitlich nachgelagerten Audiosegment beginnt und zeitlich nach einem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich des zeitlich nachgelagerten Audiosegmentes endet,
- die Dauer und Lage der verwendeten Bereiche in Abhängigkeit der vor- und nachgelagerten Audiosegmente zu bestimmen, und
- den verwendeten Bereich des zeitlich vorgelagerten Audiosegments mit dem verwendeten Bereich des zeitlich nachgelagerten Audiosegments durch Definition des Moment der Konkatenation in Abhängigkeit von Eigenschaften des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments in einem Bereich zu konkatenieren, der zeitlich unmittelbar vor dem verwendeten Bereich des zeitlich nachgelagerten Audiosegments beginnt und mit diesem endet.

17. Vorrichtung nach Anspruch 16, dadurch gekennzeichnet, daß die Konkatenationeinrichtung (111) umfaßt:
- Einrichtungen zur Konkatenation des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments mit dem verwendeten Bereich des zeitlich nachgelagerten Audiosegment, dessen verwendeter Bereich am Anfang einen statischen Laut wiedergibt, in der Umgebung der Grenzen des zuerst auftretetenden Soloartikulationsbereichs des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments,
- Einrichtungen zur Bearbeitung eines zeitlich hinteren Bereiches des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und eines zeitlich vorderen Bereiches des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen, und

- Einrichtungen zur überlappenden Addition der beiden Bereiche in einem von den zu konkatenierenden Audiosegmenten abhängenden Überlappungsbereich (Crossfade), wobei die Übergangsfunktionen und die Länge eines Überlappungsbereiches der beiden Bereiche in Abhängigkeit der zu synthetisierenden akustischen Daten bestimmt werden.

18. Vorrichtung nach Anspruch 16 oder 17, dadurch gekennzeichnet, daß die Konkatenationeinrichtung (111) umfaßt:
- Einrichtungen zur Konkatenation des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments mit dem verwendeten Bereich des zeitlich nachgelagerten Audiosegment, dessen verwendeter Bereich am Anfang einen dynamischen Laut wiedergibt, zeitlich unmittelbar vor dem verwendeten Bereich des zeitlich nachgelagerten Audiosegments,
- Einrichtungen zur Bearbeitung eines zeitlich hinteren Bereiches des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und eines zeitlich vorderen Bereiches des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen, wobei die Übergangsfunktionen in Abhängigkeit der zu synthetisierenden akustischen Daten bestimmt werden, und
- Einrichtungen zur nicht überlappenden Verbindung der Audiosegmente.

19. Vorrichtung nach einem der Ansprüche 16 bis 18, dadurch gekennzeichnet, daß die Datenbank (107) Audiosegmente enthält oder die vorgeschaltete Syntheseeinrichtung (108) Audiosegmente liefert, die Bereiche enthalten, die zu Beginn einen Laut oder einen Teil der konkatenierten Lautfolge am Anfang der konkatenierten Lautfolge wiedergibt.

20. Vorrichtung nach einem der Ansprüche 16 bis 19, dadurch gekennzeichnet, daß die Datenbank (107) Audiosegmente enthält oder die vorgeschaltete Syntheseeinrichtung (108) Audiosegmente liefert, die Bereiche enthalten, deren Ende einen Laut oder einen Teil der konkatenierten Lautfolge am Ende der konkatenierten Lautfolge wiedergibt.

21. Vorrichtung nach einem der Ansprüche 16 bis 19, dadurch gekennzeichnet, daß die Datenbank (107) eine Gruppe von Audiosegmenten enthält oder die vorgeschaltete Syntheseeinrichtung (108) Audiosegmente liefert, die Bereiche enthalten, deren Anfänge jeweils nur einen statischen Laut wiedergeben.

22. Vorrichtung nach einem der Ansprüche 16 bis 21, dadurch gekennzeichnet, daß die Konkatenationseinrichtung (111) umfaßt:
- Einrichtungen zur Erzeugung weiterer Audiosegmente durch Konkatenation von Bereichen von Audiosegmenten, wobei die Anfänge der Bereiche jeweils einen statischen Laut

wiedergeben, jeweils mit einem Bereich eines zeitlich nachgelagerten Audiosegment, dessen verwendeter Bereich am Anfang einen dynamischen Laut wiedergibt, und
- eine Einrichtung, die die weiteren Audiosegmente der Datenbank (107) oder der Auswahleinrichtung (105) zuführt.

23. Vorrichtung nach einem der Ansprüche 16 bis 22, dadurch gekennzeichnet, daß die Auswahleinrichtung (105) geeignet ist, bei der Auswahl der Audiosegmentbereiche aus der Datenbank (107) oder der vorgeschalteten Syntheseeinrichtung (108), die Audiosegmentbereiche auszuwählen, die jeweils die meisten aufeinanderfolgenden Teile der konkatenierten Laute der konkatenierten Lautfolge wiedergeben.

24. Vorrichtung nach einem der Ansprüche 16 bis 23, dadurch gekennzeichnet, daß die Konkatenationseinrichtung (111) Einrichtungen zur Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der konkatenierten Lautfolge aufweist, wobei die Funktionen u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums betreffen können.

25. Vorrichtung nach einem der Ansprüche 16 bis 24, dadurch gekennzeichnet, daß
- die Konkatenationseinrichtung (111) Einrichtungen zur Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in einem den Moment der Konkatenation umfassenden Bereich aufweist, wobei die Funktionen u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums betreffen können.

26. Vorrichtung nach einem der Ansprüche 16 bis 25, dadurch gekennzeichnet, daß
- die Konkatenationseinrichtung (111) Einrichtungen zur Auswahl des Momentes der Konkatenation bei einer Stelle in den verwendeten Bereichen des zeitlich vorgelagerten und/oder des zeitlich nachgelagerten Audiosegments aufweist, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen, wobei die Eigenschaften u.a. Nullstellen, Amplitudenwerte, Steigungen, Ableitungen beliebigen Grades, Spektren, Tonhöhen, Amplitudenwerte in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften sein können.

27. Vorrichtung nach einem der Ansprüche 16 bis 26, dadurch gekennzeichnet, daß
- die Auswahleinrichtung (105) Einrichtungen zur Implementation heuristischen Wissens umfaßt, das die Auswahl der verwendeten Bereiche der einzelnen Audiosegmente, deren Bearbeitung, deren Variation sowie deren Konkatenation betrifft.

28. Vorrichtung nach einem der Ansprüche 16 bis 27, dadurch gekennzeichnet, daß
- die Datenbank (107) Audiosegmente enthält oder die vorgeschaltete Syntheseeinrichtung (108) Audiosegmente liefert, die Bereiche enthalten, die jeweils wenigstens einen Teil eines Lautes bzw. Phons, einen Laut bzw. ein Phon, Teile von Lautfolgen bzw. Polyphonen oder Lautfolgen bzw. Polyphone wiedergeben.

29. Vorrichtung nach einem der Ansprüche 17 bis 28, dadurch gekennzeichnet, daß
- die Datenbank (107) Audiosegmente enthält oder die vorgeschaltete Syntheseeinrichtung (108) Audiosegment liefert, bei denen ein statischer Laut einem statischen Phon entspricht und Vokale, Diphtonge, Liquide, Vibranten, Frikative und Nasele umfaßt.

30. Vorrichtung nach einem der Ansprüche 18 bis 29, dadurch gekennzeichnet, daß
- die Datenbank (107) Audiosegmente enthält oder die vorgeschaltete Syntheseeinrichtung (108) Audiosegmente liefert, bei denen ein dynamischer Laut einem dynamischen Phon entspricht und Plosive, Affrikate, Glottalstops und geschlagene Laute umfaßt.

31. Vorrichtung nach einem der Ansprüche 16 bis 30, dadurch gekennzeichnet, daß
- die Konkatenationseinrichtung (111) geeignet ist, um durch Konkatenation von Audiosegmenten synthetisierte Sprachdaten zu erzeugen.

32. Vorrichtung nach einem der Ansprüche 16 bis 31, dadurch gekennzeichnet, daß
- Einrichtungen (117) zur Umwandlung der synthetisierten akustischen Daten in akustische Signale und/oder Sprachsignale vorhanden sind.

33. Datenträger, der ein Computerprogramm zur koartikulationsgerechten Konkatenation von Audiosegmenten enthält, um synthetisierte akustische Daten zu erzeugen, die eine Folge konkatenierter Laute wiedergeben, mit folgenden Schritten:
- Auswahl von wenigstens zwei Audiosegmenten, die Bereiche enthalten, die jeweils einen Teil eines Lautes oder einen Teil der Folge konkatenierter Laute wiedergeben, gekennzeichnet durch die Schritte:
- Festlegen eines zu verwendenden Bereiches eines zeitlich vorgelagerten Audiosegments,
- Festlegen eines zu verwendenden Bereiches eines zeitlich nachgelagerten Audiosegments, der zeitlich mit dem zeitlich nachgelagerten Audiosegment beginnt und mit dem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich des zeitlich nachgelagerten Audiosegments endet,

- wobei die Dauer und Lage der zu verwendenden Bereiche in Abhängigkeit der vor- und nachgelagerten Audiosegmente bestimmt wird, und
- Konkatenieren des festgelegten Bereiches des zeitlich vorgelagerten Audiosegments mit dem festgelegten Bereich des zeitlich nachgelagerten Audiosegments, indem der Moment der Konkatenation in Abhängigkeit von Eigenschaften des zu verwendenden Bereiches des zeitlich nachgelagerten Audiosegments in einen Bereich gelegt wird, der zeitlich unmittelbar vor dem zu verwendenden Bereich des nachgelagerten Audiosegments beginnt und mit diesem endet.

34. Datenträger nach Anspruch 33, dadurch gekennzeichnet, daß das Computerprogramm den Moment der Konkatenation des verwendeten Bereiches des zweiten Audiosegmentes mit dem verwendeten Bereich des ersten Audiosegment so wählt, daß
- der Moment der Konkatenation in einen Bereich gelegt wird, der in der Umgebung der Grenzen des zuerst verwendeten Soloartikulationsbereichs des zeitlich nachgelagerten Audiosegments liegt, wenn dessen verwendeter Bereich am Anfang einen statischen Laut wiedergibt, und
- ein zeitlich hinterer Bereich des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und überlappend addiert werden (Crossfade), wobei Übergangsfunktionen und die Länge eines Überlappungsbereichs der beiden Bereiche in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt wird.

35. Datenträger nach Anspruch 33 oder 34, dadurch gekennzeichnet, daß das Computerprogramm den Moment der Konkatenation des verwendeten Bereiches des zweiten Audiosegmentes mit dem verwendeten Bereich des ersten Audiosegmentes so wählt, daß
- der Moment der Konkatenation in einen Bereich gelegt wird, der zeitlich unmittelbar vor dem verwendeten Bereich des zeitlich nachgelagerten Audiosegments liegt, wenn dessen verwendeter Bereich am Anfang einen dynamischen Laut wiedergibt, und
- ein zeitlich hinterer Bereich des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und nicht überlappend verbunden werden (Hardfade), wobei die Übergangsfunktionen in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt werden.

36. Datenträger nach einem der Ansprüche 33 bis 35, dadurch gekennzeichnet, daß das Computerprogramm für einen Laut oder einen Teil der Folge konkatenierter Laute am

Anfang der konkatenierten Lautfolge einen Bereich eines Audiosegments auswählt, dessen Anfang die Eigenschaften des Anfangs der konkatenierten Lautfolge wiedergibt.

37. Datenträger nach einem der Ansprüche 33 bis 36, dadurch gekennzeichnet, daß das Computerprogramm für einen Laut oder einen Teil der Folge konkatenierter Laute am Ende der konkatenierten Lautfolge einen Bereich eines Audiosegments auswählt, dessen Ende die Eigenschaften des Endes der konkatenierten Lautfolge wiedergibt.

38. Datenträger nach einem der Ansprüche 33 bis 37, dadurch gekennzeichnet, daß das Computerprogramm eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der Lautfolge durchführt, wobei die Funktionen u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums betreffen können.

39. Datenträger nach einem der Ansprüche 33 bis 38, dadurch gekennzeichnet, daß das Computerprogramm für den zeitlich nachgelagerten Audiosegmentbereich einen Audiosegmentbereich wählt, der die größte Anzahl aufeinanderfolgender Teile der konkatenierter Laute der Lautfolge wiedergibt, um bei der Erzeugung der synthetisierten akustischen Daten die kleinste Anzahl von Audiosegmentbereichen zu verwenden.

40. Datenträger nach einem der Ansprüche 33 bis 39, dadurch gekennzeichnet, daß das Computerprogramm eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in einem Bereich durchführt, in dem der Moment der Konkatenation liegt, wobei die Funktionen u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums betreffen können.

41. Datenträger nach einem der Ansprüche 33 bis 40, dadurch gekennzeichnet, daß Computerprogramm den Moment der Konkatenation bei einer Stelle in den verwendeten Bereichen des ersten und/oder des zweiten Audiosegmentes festlegt, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen, wobei die Eigenschaften u.a. Nullstellen, Amplitudenwerte, Steigungen, Ableitungen beliebigen Grades, Spektren, Tonhöhen, Amplitudenwerte in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften sein können.

42. Datenträger nach einem der Ansprüche 33 bis 41, dadurch gekennzeichnet, daß das Computerprogramm eine Implementation von heuristischem Wissen durchführt, das die

Auswahl der verwendeten Bereiche der einzelnen Audiosegmente, deren Bearbeitung, deren Variation sowie deren Konkatenation betrifft.

43. Datenträger nach einem der Ansprüche 33 bis 42, dadurch gekennzeichnet, daß das Computerprogramm zur Erzeugung synthetisierter Sprachdaten geeignet ist, wobei die Laute Phone sind.

44. Datenträger nach einem der Ansprüche 34 bis 42, dadurch gekennzeichnet, daß das Computerprogramm zur Erzeugung statischer Laute geeignet ist, wobei die statischen Laute, Vokale, Diphtonge, Liquide, Vibranten, Frikative und Nasale umfassen.

45. Datenträger nach einem der Ansprüche 35 bis 44, dadurch gekennzeichnet, daß das Computerprogramm zur Erzeugung dynamischer Laute geeignet ist, wobei die dynamischen Laute Plosive, Affrikate, Glottalstops und geschlagene Laute

46. Datenträger nach einem der Ansprüche 33 bis 45, dadurch gekennzeichnet, daß das Computerprogramm die synthetisierten akustischen Daten in akustische umwandelbare Daten und/oder Sprachsignale umwandelt.

47. Synthetisierte Sprachsignale, die aus einer Folge von Lauten bzw. Phonen bestehen, wobei die Sprachsignale erzeugt werden, indem:
- wenigstens zwei die Laute bzw. Phone wiedergebende Audiosegmente ausgewählt werden, und
- die Audiosegmente durch eine koartikulationsgerechte Konkatenation verkettet werden, wobei
- ein zu verwendender Bereich eines zeitlich vorgelagerten Audiosegments festgelegt wird,
- ein zu verwendender Bereich eines zeitlich nachgelagerten Audiosegments festgelegt wird, der zeitlich mit dem zeitlich nachgelagerten Audiosegment beginnt und mit dem auf den zuerst verwendeten Soloartikulationsbereich folgenden Koartikulationsbereich des zeitlich nachgelagerten Audiosegments endet,
- wobei die Dauer und Lage der zu verwendenden Bereiche in Abhängigkeit der Audiosegmente bestimmt wird, und
- die verwendeten Bereiche der Audiosegmente koartikulationsgerecht konkateniert werden, indem der Moment der Konkatenation in Abhängigkeit von Eigenschaften des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments in einem Bereich festge-

legt wird, der unmittelbar vor dem zu verwendenden Bereich des zeitlich nachgelagerten Audiosegments beginnt und mit diesem endet.

48. Synthetisierte Sprachsignale nach Anspruch 47, dadurch gekennzeichnet, daß die Sprachsignale erzeugt werden, indem
- die Audiosegmente zu einem Moment konkateniert werden, der in der Umgebung der Grenzen des zuerst auftretenden Soloartikulationsbereichs des verwendeten Bereiches des zeitlich nachgelagerten Audiosegmentes liegt, wenn der Anfang dieses Bereiches einen statischen Laut bzw. ein statisches Phon wiedergibt , wobei ein statisches Phon ein Vokal, ein Diphtong, ein Liquid, ein Frikativ, ein Vibrant oder ein Nasal ist, und
- ein zeitlich hinterer Bereich des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet und beide Bereiche überlappend addiert werden (Crossfade), wobei die Übergangsfunktionen und die Länge eines Überlappungsbereichs beiden Bereiche in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt werden.

49. Synthetisierte Sprachsignale nach Anspruch 47 oder 48, dadurch gekennzeichnet, daß die Sprachsignale erzeugt werden, indem
- die Audiosegmente zu einem Moment konkateniert werden, der zeitlich unmittelbar vor dem verwendeten Bereich des zeitlich nachgelagerten Audiosegmentes liegt, wenn der Anfang dieses Bereiches einen dynamischen Laut bzw. ein dynamisches Phon wiedergibt, wobei ein dynamisches Phon ein Plosiv, ein Affrikat, ein Glottalstop oder ein geschlagener Laut ist, und
- ein zeitlich hinterer Bereich des verwendeten Bereiches des zeitlich vorgelagerten Audiosegments und ein zeitlich vorderer Bereich des verwendeten Bereiches des zeitlich nachgelagerten Audiosegments mit geeigneten Übergangsfunktionen bearbeitet werden und nicht überlappend verbunden werden (Hardfade) wobei die Übergangsfunktionen in Abhängigkeit der zu konkatenierenden Audiosegmente bestimmt werden.

50. Synthetisierte Sprachsignale nach einem der Ansprüche 47 bis 49, dadurch gekennzeichnet, daß
- der erste Laut bzw. das erste Phon oder ein Teil der ersten Lautfolge bzw. des ersten Polyphons in der Folge durch ein Audiosegment erzeugt wird, dessen verwendeter Bereich am Anfang die Eigenschaften des Anfangs der Folge wiedergibt.

51. Synthetisierte Sprachsignale nach einem der Ansprüche 47 bis 50, dadurch gekennzeichnet, daß
- der letzte Laut bzw. das letzte Phon oder ein Teil der letzten Lautfolge bzw. des letzten Polyphon in der Folge durch ein Audiosegment erzeugt wird, dessen verwendeter Bereich am Ende die Eigenschaften des Endes der Folge wiedergibt.

52. Synthetisierte Sprachsignale nach einem der Ansprüche 47 bis 51, dadurch gekennzeichnet, daß
- die Sprachsignale erzeugt werden, indem nachgelagerte mit der Wiedergabe eines dynamischen Lautes bzw. Phons beginnenden Bereiche von Audiosegmenten mit vorgelagerten mit der Wiedergabe eines statischen Lautes bzw. Phons beginnende Bereichen von Audiosegmenten konkateniert werden.

53. Synthetisierte Sprachsignale nach einem der Ansprüche 47 bis 52, dadurch gekennzeichnet, daß
- die Audiosegmentbereiche ausgewählt werden, die die meisten Teile von Lauten bzw. Phonen der Folge wiedergeben, um bei der Erzeugung der Sprachsignale die minimale Anzahl von Audiosegmentbereichen zu verwenden.

54. Synthetisierte Sprachsignale nach einem der Ansprüche 47 bis 53, dadurch gekennzeichnet, daß
- die Sprachsignale durch Konkatenation der verwendeten Bereiche von Audiosegmenten erzeugt werden, die mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der Lautfolge bzw. Phonfolge bearbeitet werden, wobei die Funktionen u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums betreffen können.

55. Synthetisierte Sprachsignale einem der Ansprüche 47 bis 54, dadurch gekennzeichnet, daß
- die Sprachsignale durch Konkatenation der verwendeten Bereiche von Audiosegmenten erzeugt werden, die mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der Lautfolge bzw. Phonfolge in einem Bereich bearbeitet werden, in dem der Moment der Konkatenation liegt, wobei die Funktionen u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums betreffen können.

56. Synthetisierte Sprachsignale einem der Ansprüche 47 bis 55, dadurch gekennzeichnet, daß der Moment der Konkatenation bei einer Stelle in den verwendeten Bereichen des vorgelagerten und/oder des nachgelagerten Audiosegmentes liegt, an denen die beiden

verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen, wobei diese Eigenschaften u.a. Nullstellen, Amplitudenwerte, Steigungen, Ableitungen beliebigen Grades, Spektren, Tonhöhen, Amplitudenwerte in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften sein können.

57. Synthetisierte Sprachsignale nach einem der Ansprüche 47 bis 56, dadurch gekennzeichnet, daß die Sprachsignale geeignet sind, in akustische Signale umgewandelt zu werden.

58. Akustischer, optischer, magnetischer oder elektrischer Datenspeicher, der Audiosegmente enthält, um durch eine Konkatenation von verwendeten Bereichen der Audiosegmente unter Verwendung des Verfahrens nach Anspruch 1 oder der Vorrichtung nach Anspruch 16 oder des Datenträgers nach Anspruch 33 synthetisierte akustische Daten zu erzeugen.

59. Datenspeicher nach Anspruch 58, dadurch gekennzeichnet, daß eine Gruppe der Audiosegmente Laute bzw. Phone oder Teile von Lauten bzw. Phonen wiedergeben.

60. Datenspeicher nach Anspruch 58 oder 59, dadurch gekennzeichnet, daß eine Gruppe der Audiosegmente Lautfolgen oder Teile von Lautfolgen bzw. Polyphone oder Teile von Polyphonen wiedergeben.

61. Datenspeicher nach einem der Ansprüche 58 bis 60, dadurch gekennzeichnet, daß eine Gruppe von Audiosegmenten zur Verfügung gestellt wird, deren verwendete Bereiche mit einem statischen Laut bzw. Phon beginnen, wobei die statischen Phone Vokale, Diphtonge, Liquide, Frikative, Vibranten und Nasale umfassen.

62. Datenspeicher nach einem der Ansprüche 58 bis 61, dadurch gekennzeichnet, daß Audiosegmente zur Verfügung gestellt werden, die geeignet sind in akustische Signale umgewandelt zu werden.

63. Datenspeicher nach einem der Ansprüche 58 bis 62, der zusätzlich Informationen enthält, um eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente mit Hilfe geeigneter Funktionen in Abhängigkeit von Eigenschaften der zu synthetisierenden akustischen Daten durchzuführen, wobei die Funktionen u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums betreffen können.

64. Datenspeicher nach einem der Ansprüche 58 bis 63, der zusätzlich Informationen enthält, die eine Bearbeitung der verwendeten Bereiche einzelner Audiosegmente und mit Hilfe geeigneter Funktionen in einem Bereich betreffen, in dem der Moment der Konkatenation liegt, wobei die Funktionen u.a. eine Veränderung der Frequenz, der Dauer, der Amplitude oder des Spektrums betreffen können.

65. Datenspeicher nach einem der Ansprüche 58 bis 64, der zusätzlich verkettete Audiosegmente zur Verfügung stellt, deren Moment der Konkatenation bei einer Stelle der verwendeten Bereiche des zeitlich vorgelagerten und/oder des zeitlich nachgelagerten Audiosegmentes liegt, an denen die beiden verwendeten Bereiche hinsichtlich einer oder mehrerer geeigneter Eigenschaften übereinstimmen, wobei die Eigenschaften u.a. Nullstellen, Amplitudenwerte, Steigungen, Ableitungen beliebigen Grades, Spektren, Tonhöhen, Amplitudenwerte in einem Frequenzbereich, Lautstärke, Sprachstil, Sprachemotion, oder andere im Lautklassifizierungsschema betrachtete Eigenschaften sein können.

66. Datenspeicher nach einem der Ansprüche 58 bis 65, der zusätzlich Informationen in Form von heuristischem Wissen enthält, die die Auswahl der verwendeten Bereiche der einzelnen Audiosegmente, deren Bearbeitung, deren Variation sowie deren Konkatenation betreffen.

67. Tonträger, der Daten enthält, die zumindest teilweise synthetisierte akustische Daten sind, die
- mit dem Verfahren nach Anspruch 1, oder
- mit der Vorrichtung nach Anspruch 16, oder
- unter Verwendung des Datenträgers nach Anspruch 33 , oder
- unter Verwendung des Datenspeichers nach Anspruch 58 erzeugt wurden, oder
- die Sprachsignale nach Anspruch 47 sind.

68. Tonträger nach Anspruch 67, dadurch gekennzeichnet, daß die synthetisierten akustischen Daten synthetisierte Sprachdaten sind.

09/763149
5060
Translation

PATENT COOPERATION TREATY

09/763149

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference EP-82 972/PC	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/EP99/06081	International filing date (*day/month/year*) 19 August 1999 (19.08.99)	Priority date (*day/month/year*) 19 August 1998 (19.08.98)
International Patent Classification (IPC) or national classification and IPC G10L 13/06		
Applicant BUSKIES, Christoph		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 6 sheets, including this cover sheet.

☒ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of 14 sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☒ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☒ Certain defects in the international application
- VIII ☒ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
17 March 2000 (17.03.00)	02 October 2000 (02.10.2000)
Name and mailing address of the IPEA/EP Facsimile No.	Authorized officer Telephone No.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/EP99/06081

I. Basis of the report

1. This report has been drawn on the basis of (*Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

☐ the international application as originally filed.

☒ the description, pages 1-22, as originally filed,
pages _____, filed with the demand,
pages _____, filed with the letter of _____,
pages _____, filed with the letter of _____.

☒ the claims, Nos. _____, as originally filed,
Nos. _____, as amended under Article 19,
Nos. _____, filed with the demand,
Nos. 1-68, filed with the letter of 24 August 2000 (24.08.2000),
Nos. _____, filed with the letter of _____.

☒ the drawings, sheets/fig 1/13-13/13, as originally filed,
sheets/fig _____, filed with the demand,
sheets/fig _____, filed with the letter of _____,
sheets/fig _____, filed with the letter of _____.

2. The amendments have resulted in the cancellation of:

☐ the description, pages _____
☐ the claims, Nos. _____
☐ the drawings, sheets/fig _____

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).

4. Additional observations, if necessary:

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 58-66

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international preliminary examination *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. 58-66 are so unclear that no meaningful opinion could be formed *(specify)*:

See separate sheet

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed.

☐ no international search report has been established for said claims Nos. _____.

Supplemental Box
(To be used when the space in any of the preceding boxes is not sufficient)

Continuation of: BOX III

1. Independent Claim 58 and Claims 59-66, which are dependent thereon, lay claim to a data memory containing audio segments. However, these audio segments are characterised only by the result to be achieved, that is, they must be suitable for generating synthesised acoustic data using the method as per Claim 1, the device as per Claim 16 or the data carrier as per Claim 33.

   Since this does not define any features of the audio segments in the claim itself (PCT Rule 6.3(a)), the claim is not clear (PCT Article 6; see also PCT Examination Guidelines, Chapter III-4.7).

2. Consequently, Claims 58-66 were not considered when establishing the examination report (PCT Article 34(4)(a)(ii)).

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Claims	1-57, 67-68	YES
	Claims		NO
Inventive step (IS)	Claims	1-57, 67-68	YES
	Claims		NO
Industrial applicability (IA)	Claims	1-57, 67-68	YES
	Claims		NO

2. Citations and explanations

1. The present application concerns speech synthesis, in particular data-based systems in which a neat concatenation of the individual speech pattern elements in the memory has proven critical to the synthesis quality attained.

The closest prior art is disclosed in document "A TtS system for the Greek language based on the concatenation of formant coded segments", Yourgalis et al., Speech Communication 19(1996), pages 21-38. That document shows that the coarticulation effect can be allowed for by corresponding simulation calculations and that other concatenation modes can be used, depending on phoneme type, which in their respective classes can yield better results than the other methods available for selection.

The technical problem addressed by the present application is that of finding an alternative to the approaches proposed in the closest prior art.

For that purpose, an entirely data-based approach is used to tackle the coarticulation problem. The section to be used of the audio segment that comes

later in time (and which is to be attached after the former result) ends with the coarticulation section of this later audio segment that follows the solo articulation section that was first used.

Furthermore, the moment (in time) and not the mode of concatenation is determined in the present application as a function of the properties of adjacent sections, in a situation-dependent way.

This is claimed in the independent method, device and program data carrier Claims 1, 16 and 33.

The use of the steps listed in Claim 1 is therefore neither disclosed nor suggested by D1. None of the remaining international search report citations contains any hints that would lead a person skilled in the art to replace the prior art in D1 by the features of Claim 1.

Claim 1 is therefore novel and involves an inventive step (PCT Article 33(2) and (3)).

This also applies to independent device Claim 16 and independent program data carrier Claim 33, as well as to independent product Claim 47 and independent product data carrier Claim 67, insofar as these do not refer back to the unclear Claim 58. Furthermore, this also applies to dependent Claims 2-15, 17-32, 34-46, 48-57 and 68 (insofar as Claim 67 does not refer back to Claim 58), which refer to these independent claims without creating any contradiction with the independent claims to which they refer.

VII. Certain defects in the international application

The following defects in the form or contents of the international application have been noted:

Contrary to PCT Rule 5.1(a)(ii), the description does not cite document D1 and does not indicate the relevant prior art disclosed therein.

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

1. At least one word is missing in the description, page 9, in the sentence that extends from line 32 to line 37. This makes the sentence incomprehensible. The Examiner assumes that the word "determined" in line 37 should have been replaced by "is determined".

2. There is discrepancy between page 10, lines 21-30 and page 12, lines 14-20 on the one hand, and page 17, lines 24-30 and page 18, lines 25-34 on the other hand. In the former passages, audio segment sections which begin with the reproduction of a dynamic sound are preferred, while in the latter passages exactly the opposite is preferred, that is those beginning with a static sound. The reader cannot know in which of the two cases the advantage of a reduced outlay (which is mentioned in the two cases) is actually obtained. This lack of clarity should be eliminated.

3. The claim numbers used in the description do not match the actual claim numbers.

Claims

1. A method for the co-articulation-specific concatenation of audio segments, in order to generate synthesised acoustical data which reproduces a sequence of concatenated sounds/phones, comprising the following steps:
- selection of at least two audio segments which contain bands, each of which reproducing a portion of a sound/phone or a portion of a sound/phone sequence,
characterised by the steps of:
- establishing a band to be used of an earlier audio segment;
- establishing a band to be used of a later audio segment, which begins immediately before the band to be used of the later audio segment and ends with the co-articulation band of the later audio segment which follows the initially used solo articulation band;
- with the duration and position of the bands to be used being determined as a function of the earlier and later audio segments; and
- concatenating the established band of the earlier audio segment with the established band of the later audio segment, in that the instance of concatenation, as a function of properties of the used band of the later audio segment, is set in its established band.

2. The method according to Claim 1, characterised in that
- the instance of concatenation is set in a band which lies in the vicinity of the boundaries of the initially to be used solo articulation band of the later audio segment, if the band of same to be used reproduces a static sound/phone at the beginning; and
- a downstream portion of the band to be used of the earlier audio segment and an upstream portion of the band to be used of the later audio segment are processed by means of suitable

transfer functions and added in an overlapping manner (cross fade), with the transfer functions and the length of an overlapping portion of the two bands being determined depending on the audio segments to be concatenated.

3. The method according to Claim 1, characterised in that
- the instance of concatenation is set in a band which lies immediately before the band to be used of the later audio segment, if the used band of same reproduces a dynamic sound/phone at the beginning; and
- a downstream portion of the band to be used of the earlier audio segment and an upstream portion of the band to be used of the later audio segment are processed by means of suitable transfer functions and joined in a non-overlapping manner (hard fade), with the transfer functions being determined depending on the acoustical data to be synthesised.

4. The method according to one of Claims 1 to 3, characterised in that for a sound/phone or a portion of the sequence of concatenated sounds/phones at the start of the concatenated sound/phone sequence a band of an audio segment is selected so that the start of the band reproduces the properties of the start of the concatenated sound/phone sequence.

5. The method according to one of Claims 1 to 4, characterised in that for a sound/phone or a portion of the sequence of concatenated sounds/phones at the end of the concatenated sound/phone sequence a band of an audio segment is selected so that the end of the band reproduces the properties of the end of the concatenated sound/phone sequence.

6. The method according to one of Claims 1 to 5, characterised in that the voice data to the synthesised is combined in groups, each of which being described by an individual audio segment.

7. The method according to one of Claims 1 to 6, characterised in that an audio segment is selected for the later audio segment band, which reproduces the highest number of successive portions of the sounds/phones of the sound/phone sequence, in order to use the smallest number of audio segment bands in the generation of the synthesised acoustical data.

8. The method according to one of Claims 1 to 7, characterised in that a processing of the used bands of individual audio segments is carried out by means of suitable functions depending on properties of the concatenated sound/phone sequence, with these properties involving i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

9. The method according to one of Claims 1 to 8, characterised in that a processing of the used bands of individual audio segments is carried out by means of suitable functions in a band, in which the instance of concatenation lies. This can include i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

10. The method according to one of Claims 1 to 9, characterised in that the instance of concatenation is set in places of the bands to be used of the earlier and/or later audio segment, in which the two used bands are in agreement with respect to one or several suitable properties, with these properties including i.a.: zero point, amplitude value, gradient, derivative of any degree, spectrum, tone level, amplitude value in a frequency band, volume, style of speech, emotion of speech, or other properties covered in the phone classification scheme.

11. The method according to one of Claims 1 to 10, characterised in that

- the selection of the used bands of individual audio segments, their processing, their variation, as well as their concatenation are additionally carried out with the application of heuristic knowledge which is obtained by an additionally carried out heuristic method.

12. The method according to one of Claims 1 to 11, characterised in that
- the acoustical data to be synthesised is voice data, and the sounds are phones;
- the static phones include vowels, diphtongs, liquids, vibrants, fricatives and nasals; and
- the dynamic phones include plosives, affricates, glottal stops, and click sounds.

13. The method according to one of Claims 1 to 12, characterised in that
- a conversion of the synthesised acoustical data to acoustical signals and/or voice signals is carried out.

14. A device for the co-articulation-specific concatenation of audio segments, in order to generate synthesised acoustical data which reproduces a sequence of phones, comprising:
- a database in which audio segments are stored, each of which reproducing portion of a phone or portions of a sequence of (concatenated) phones;
- and/or any upstream synthesis means (not part of this invention) which supplies audio segments;
- a means for the selection of at least two audio segments from the database and/or the upstream synthesis means; and
- a means for the concatenation of audio segments, characterised in that the concatenation means is suited for
- defining a band to be used of an earlier audio segment;
- defining a portion to be used of a later audio segment in a band which starts with the later audio segment and ends after

a co-articulation band of the later audio segment, which follows after the initially used solo articulation band;
- determining the duration and position of the used bands depending on the earlier and later audio segments; and
- concatenating the used band of the earlier audio segment with the used band of the later audio segment by defining the instance of concatenation as a function of properties of the used band of the later audio segment in a band which starts immediately before the used band of the later audio segment and ends with the co-articulation band which follows after the initially used solo articulation band after of the later audio segment.

15. The device according to Claim 14, characterised in that the concatenation means comprises:
- means for the concatenation of the used band of the earlier audio segment with the used band of the later audio segment, whose used band reproduces a static phone at the beginning in the vicinity of the boundaries of the initially occurring solo articulation band of the used band of the later audio segment;
- means for processing a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment by suitable transfer functions; and
- means for the overlapping addition of the two bands in an overlapping portion (cross fade), which depends on the audio segments to be concatenated, with the transfer functions and the length of an overlapping portion of the two bands being determined depending on the acoustical data to be synthesised.

16. The device according to Claim 14, characterised in that the concatenation means comprises:
- means for the concatenation of the used band of the earlier audio segment with the used band of the later audio segment,

whose used band reproduces a dynamic phone at the beginning, immediately before the used band of the later audio segment;
- means for processing a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment by suitable transfer functions, with the transfer functions being determined depending on the acoustical data to be synthesised; and
- means for the non-overlapping joining of the two audio segments.

17. The device according to one of Claims 14 to 16, characterised in that the database includes audio segments or the upstream synthesis means supplies audio segments which comprise bands which at the start reproduce a phone or a portion of the concatenated phone sequence at the start of the concatenated phone sequence.

18. The device according to one of Claims 14 to 17, characterised in that the database includes audio segments or the upstream synthesis means supplies audio segments which comprise bands, whose ends reproduce a phone or a portion of the concatenated phone sequence at the end of the concatenated phone sequence.

19. The device according to one of Claims 14 to 18, characterised in that the database includes a group of audio segments or the upstream synthesis means supplies audio segments which comprise bands, whose starts each reproduce only a static phone.

20. The device according to one of Claims 14 to 19, characterised in that the concatenation means comprises:
- means for the generation of further audio segments by concatenation of audio segments, with the starts of the bands each reproducing a static phone, each with a band of a later

audio segment whose used band reproduces a dynamic phone at the start, and
- a means which supplies the further audio segments to the database or the selection means.

21. The device according to one of Claims 14 to 20, characterised in that, in the selection of the audio segment bands from the database or the upstream synthesis means, the selection means is suited to select the audio segments which reproduce the greatest number of successive portions of concatenated phones of the concatenated phone sequence.

22. The device according to one of Claims 14 to 21, characterised in that the concatenation means comprises means for processing the used bands of individual audio segments with the aid of suitable functions, depending on properties of the concatenated phone sequence. Among others, this can be a modification of the frequency, the duration, the amplitude, or the spectrum.

23. The device according to one of Claims 14 to 22, characterised in that
- the concatenation means comprises means for processing the used bands of individual audio segments with the aid of suitable functions in a band including the instance of concatenation, with this function involving i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

24. The device according to one of Claims 14 to 23, characterised in that
- the concatenation means comprises means for the selection of the instance of concatenation in a place in the used bands of the earlier and/or the later audio segment, in which the two used bands are in agreement with respect to one or several suitable properties, with these properties including i.a.:

zero point, amplitude value, gradient, derivatives of any degree, spectrum, tone level, amplitude value in a frequency band, volume, style of speech, emotion of speech, or other properties covered in the phone classification scheme.

25. The device according to one of Claims 14 to 24, characterised in that
- the selection means comprises means for the implementation of heuristic knowledge which relates to the selection of the used bands of the individual audio segments, their processing, their variation, as well as their concatenation.

26. The device according to one of Claims 14 to 25, characterised in that
- the database includes audio segments or the upstream synthesis means supplies audio segments which include bands, each of which reproducing at least a portion of a sound or phone, respectively, a sound or phone, respectively, portions of phone sequences or polyphones, respectively, or sound/phone sequences or polyphones, respectively, with a static sound corresponding to a static phone and comprising vowels, diphtongs, liquids, vibrants, fricatives, and nasals; and
a dynamic sound corresponding to a dynamic phone and comprising plosives, affricates, glottal stops, and klick speech; and
- the concatenation means is suitable to generate synthesised voice data by means of the concatenation of audio segments.

27. The device according to one of Claims 14 to 26, characterised in that
- means are provided for the conversion of the synthesised acoustical data to acoustical signals and/or voice signals.

28. Synthesised voice signals which consist of a sequence of sounds or phones, respectively, with the voice signals being generated in that:

- at least two audio segments are selected which reproduce the sounds or phones, respectively; and
- the audio segments are linked by a co-articulation-specific concatenation, with
- a band to be used of an earlier audio segment being established;
- a band to be used of a later audio segment being established which starts immediately before the band to be used of the later audio segment and ends with the co-articulation band of the later audio segment, following the initially used solo articulation band;
- with the duration and position of the bands to be used being determined depending on the audio segments; and
- the used bands of the audio segments being concatenated in a co-articulation-specific manner, in that the instance of concatenation, as a function of properties of the used band of the later audio segment, is set in its established band.

29. The synthesised voice signals according to Claim 28, characterised in that the voice signals are generated in that
- the audio segments are concatenated in an instance which lies in the vicinity of the boundaries of the initially occurring solo articulation band of the used band of the later audio segment, if the start of this band reproduces a static sound or a static phone, respectively, with the static phone being a vowel, a diphtong, a liquid, a fricative, a vibrant, or a nasal; and
- a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment are processed by means of suitable transfer function and both bands are added in an overlapping manner (cross fade), with the transfer functions and the length of an overlapping portion of the two bands being determined depending on the audio segments to be concatenated.

30. The synthesised voice signals according to Claim 28, characterised in that the voice signals are generated in that
- the audio segments are concatenated in an instance which lies immediately before the used band of the later audio segment, if the start of this band reproduces a dynamic sound or phone, respectively, with the dynamic phone being a plosive, an affricate, a glottal stop, or klick speech; and
- a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment are processed by means of suitable transfer functions and both bands are joined in a non-overlapping manner (hard fade), with the transfer functions being determined depending on the audio segments to be concatenated.

31. The synthesised voice signals according to one of Claims 28 to 30, characterised in that
- the first sound or the first phone, respectively, or a portion of the first phone sequence or of the first polyphone, respectively, in the sequence is generated by an audio segment, whose used band at the start reproduces the properties of the start of the sequence.

32. The synthesised voice signals according to one of Claims 28 to 30, characterised in that
- the last sound or the last phone, respectively, or a portion of the last phone sequence or of the last polyphone, respectively, in the sequence is generated by an audio segment, whose used band at the end reproduces the properties of the end of the sequence.

33. The synthesised voice signals according to one of Claims 28 to 32, characterised in that
- the voice signals are generated in that later bands of audio segments, beginning with the reproduction of a dynamic sound or phone, respectively, are concatenated with earlier bands of